

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 November 2002 (14.11.2002)

PCT

(10) International Publication Number
WO 02/090530 A2

(51) International Patent Classification⁷: **C12N 9/00**

(21) International Application Number: **PCT/US02/01369**

(22) International Filing Date: 16 January 2002 (16.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/263,083	18 January 2001 (18.01.2001)	US
60/271,205	23 February 2001 (23.02.2001)	US
60/271,117	23 February 2001 (23.02.2001)	US
60/276,859	16 March 2001 (16.03.2001)	US
60/278,504	23 March 2001 (23.03.2001)	US
60/278,522	23 March 2001 (23.03.2001)	US
60/280,510	29 March 2001 (29.03.2001)	US
60/280,266	29 March 2001 (29.03.2001)	US

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **WALLA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **GURURAJAN, Rajagopal** [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 40633 (US). **JACKSON, Jennifer, L.** [US/US]; 4913 Deep

Creek Road, Fremont, CA 94555 (US). **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TRAN, Bao** [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). **DING, Li** [US/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/090530 A2

(54) Title: KINASES AND PHOSPHATASES

(57) Abstract: The invention provides human kinases and phosphatases (KPP) and polynucleotides which identify and encode KPP. The invention also provides expression vector, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of KPP.

KINASES AND PHOSPHATASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of kinases and phosphatases and to the use of these sequences in the diagnosis, treatment, and prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of kinases and phosphatases.

10

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response

to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

5 Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked
10 to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine
15 residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to
20 the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through
25 IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

35 **Protein Tyrosine Kinases**

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates

its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase domain as is typical with other MAP kinases (Abe, M.K. (1999) Mol. Cell. Biol. 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 5 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, 10 bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

15 The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. 20 Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification 25 of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α , β , γ , δ , and ϵ . Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype 30 caused by deletion of the yeast CKI locus, HRR250 (Fish et al., supra).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the Drosophila circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the 35 mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.

Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore 5 the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) *Science* 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) *Biochimie* 82:1123-1127). 10 HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) *J. Biol. Chem.* 273:25875-25879; Wang, Y. et al. (2001) *Biochim. Biophys. Acta* 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of 15 which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) *Proc. Natl. Acad. Sci. USA* 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus 20 (Nishiyama, Y. et al. (1999) *FEBS Lett.* 459:159-165).

Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) *EMBO J.* 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) *BioEssays* 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be 35 enriched in mammalian forebrain. This protein is associated with vesicles in both axons and

dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) *J. Neurosci.* 14:1-13).

Mitogen-Activated Protein Kinases

5 The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) *Nature* 365:781-783). There are three kinase modules comprising the MAP kinase 10 cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang,X.S. et al (1998) *Biochem. Biophys. Res. Commun.* 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun 15 N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

20 **Cyclin-Dependent Protein Kinases**

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the 25 mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation 30 of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) *EMBO J.* 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome 35 segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint

pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in 5 yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) *Science* 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) *Science* 277:1501-1505). Thus, activation of 10 Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation 15 of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) *J. Biol. Chem.* 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) *J. Biol. Chem.* 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and 25 hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a 35 C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This

C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) *J. Biol. Chem.* 273:29066-29071). DRAK1 and DRAK2 are STKs that share 5 homology with the death-associated protein kinases (DAP kinases), known to function in interferon- γ induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the 10 N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) *J. Biol. Chem.* 273:12296-12300). CD95 is a 15 member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) *Cell* 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal 20 kinase catalytic domain and a C-terminal “caspase-recruitment” domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

25 **Mitochondrial Protein Kinases**

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration- 30 dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) *Adv. Enzyme Regul.* 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory 35 enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to

a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) *Adv. Enzyme Regul.* 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes 5 in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

10 Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) *Curr. Opin. Cell. Biol.* 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular 15 signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP₂). PIP₂ is then cleaved into inositol triphosphate (IP₃) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by 20 these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit 25 acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP₂) to PI (3,4,5) P₃ (PIP₃). PIP₃ then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, 30 and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) *Biochem. J.* 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

35 An example of lipid kinase phosphorylation activity is the phosphorylation of

D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) *J. Biol. Chem.* 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP 5 regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive 10 inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., supra).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism 15 and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) *Cancer Res.* 50:1576-1579).

20 AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular 25 components consuming ATP (Zeleznikar, R.J. et al. (1995) *J. Biol. Chem.* 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be 30 treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production 35 of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and

GTP levels also control the activity of certain oncogenic proteins such as p21^{ras} known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) *Cancer Res.* 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21^{ras} and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may 5 provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buclovir (Miller, W.H. and R.L. Miller (1980) *J. Biol. Chem.* 255:7204-7207; Stenberg, K. et al. (1986) *J. Biol. Chem.* 261:2134-2139). Increasing GuK activity in infected cells 10 may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol 15 (Johansson, M. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) *Pharmacol. Ther.* 67:155-186).

20

PHOSPHATASES

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. 25 The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal 30 and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca²⁺ or Mn²⁺, for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. 35 Members of this class are composed of a homologous catalytic subunit bearing a very highly

conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

5 PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-
10 monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) *Curr. Opin. Neurobiol.* 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) *Otolaryngol. Head Neck Surg.* 121:463-468).

15 PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) *Trends Biosci.* 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen.
20 Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) *J. Virol.* 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the
25 activities of more than 30 protein kinases in vitro, and other evidence suggests that the same is true in vivo for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the I κ B kinases (reviewed in Millward et al., supra). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B
30 (Bastians, H. et al. (1999) *Mol. Biol. Cell* 10:3927-3941). PP2A is a major activity in the brain and is
35

implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

5 PP2B, or calcineurin, is a Ca^{2+} -activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of
10 transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, *supra*).

15 Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell
20 Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not
25 been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble
30 phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a
35 transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic

domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This 5 wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, supra).
10 CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and
15 Tonks, supra). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and
20 neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased
25 tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, supra).
30

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, supra). In the activated state, these kinases have been implicated
35

in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) *J. Biol. Chem.* 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) *Hum. Pathol.* 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura, M. et al. (1999) *J. Natl. Cancer Inst.* 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

LAP, an orthophosphoric monoester of the endosomal/lysosomal compartment is a housekeeping gene whose enzymatic activity has been detected in all tissues examined (Geier, C. et al. (1989) *Eur. J. Biochem.* 183:611-616). LAP-deficient mice have progressive skeletal disorder and an increased disposition toward generalized seizures (Saftig, P. et al. (1997) *J. Biol. Chem.* 272:18628-18635). LAP-deficient patients were found to have the following clinical features: intermittent vomiting, hypotonia, lethargy, opisthotonus, terminal bleeding, seizures, and death in early infancy (Online Mendelian Inheritance in Man (OMIM) *200950).

PAP, a prostate epithelium-specific differentiation antigen produced by the prostate gland, has been used to diagnose and stage prostate cancer. In prostate carcinomas, the enzymatic activity of PAP was shown to be decreased compared with normal or benign prostate hypertrophy cells (Foti, A. G. et al. (1977) *Cancer Res.* 37: 4120-4124). Two forms of PAP have been identified, secreted and intracellular. Mature secreted PAP is detected in the seminal fluid and is active as a glycosylated homodimer with a molecular weight of approximately 100-kilodalton. Intracellular PAP is found to exhibit endogenous phosphotyrosyl protein phosphatase activity and is involved in regulating prostate cell growth (Meng, T.C. and Lin, M.F. (1998) *J. Biol. Chem.* 34: 22096-22104).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) *FEBS Lett.* 419:175-180). Additionally,

synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) *Curr. Opin. Neurobiol.* 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, 5 suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) *Cell* 99:179-188).

The discovery of new kinases and phosphatases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, 10 prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of kinases and phosphatases.

15

SUMMARY OF THE INVENTION

The invention features purified polypeptides, kinases and phosphatases, referred to collectively as "KPP" and individually as "KPP-1," "KPP-2," "KPP-3," "KPP-4," "KPP-5," "KPP-6," "KPP-7," and "KPP-8." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group 20 consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the 25 invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ 30 ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-8. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID 35 NO:9-16.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 5 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention 10 provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 15 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) 20 culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected 25 from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide selected from the group consisting 30 of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the 35 polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method 5 comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe 10 specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, 15 said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide 20 complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a 25 polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an 30 immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such 35 treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected 5 from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the 10 invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as 15 an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, 20 and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of 25 treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected 30 from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one 35 test compound under suitable conditions, and b) detecting binding of the polypeptide to the test

compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide

complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

10 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

15 Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide 20 sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

25 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood 30 that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," 35 and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a

reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"KPP" refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

An "allelic variant" is an alternative form of the gene encoding KPP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding KPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a polypeptide with at least one functional characteristic of KPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding KPP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent KPP. Deliberate amino acid substitutions may be made on the basis of similarity in

polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. 5 Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic 10 molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. 15 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP 20 participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide 25 used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that 30 makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

35 The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a

specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include

5 deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.

10 Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

15 The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

20 The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense 25 molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

30 The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

35 "Complementary" describes the relationship between two single-stranded nucleic acid

sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding KPP or fragments of KPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
25	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
30	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
35	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
40	Trp	Phe, Tyr
	Tyr	His, Phe, Trp

Val

Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, 5 (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide.

10 Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

15 A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

“Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a 20 diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

25 A “fragment” is a unique portion of KPP or the polynucleotide encoding KPP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 30 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the 35 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present

embodiments.

A fragment of SEQ ID NO:9-16 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:9-16, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:9-16 is useful, for 5 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:9-16 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:9-16 and the region of SEQ ID NO:9-16 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-8 is encoded by a fragment of SEQ ID NO:9-16. A fragment of 10 SEQ ID NO:1-8 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-8. For example, a fragment of SEQ ID NO:1-8 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-8. The precise length of a fragment of SEQ ID NO:1-8 and the region of SEQ ID NO:1-8 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the 15 fragment.

A “full length” polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, interchangeably, sequence identity, between 20 two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and 25 therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in 30 Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and “diagonals saved”=4. The “weighted” residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polynucleotide sequences.

35 Alternatively, a suite of commonly used and freely available sequence comparison algorithms

is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at

<http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence

5 analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST
10 programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

15 *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

20 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous
25 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes
30 in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a
35 standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative

substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

10 Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

15 *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

20 Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment 25 length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

30 The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific 35 hybridization is an indication that two nucleic acid sequences share a high degree of complementarity.

Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive 5 conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

10 Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and 15 conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, 20 for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic 25 solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term “hybridization complex” refers to a complex formed between two nucleic acid 30 sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_{0t} or R_{0t} analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

35 The words “insertion” and “addition” refer to changes in an amino acid or nucleotide

sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect 5 cellular and systemic defense systems.

An “immunogenic fragment” is a polypeptide or oligopeptide fragment of KPP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term “immunogenic fragment” also includes any polypeptide or oligopeptide fragment of KPP which is useful in any of the antibody production methods disclosed herein or known in the

10 art.

The term “microarray” refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms “element” and “array element” refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

15 The term “modulate” refers to a change in the activity of KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.

20 The phrases “nucleic acid” and “nucleic acid sequence” refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the 25 antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

“Operably linked” refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding 25 sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

30 “Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNA 30 preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

35 “Post-translational modification” of an KPP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of KPP.

“Probe” refers to nucleic acid sequences encoding KPP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

5 “Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous 10 nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

15 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs 20 can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 25 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from 30 megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection 35 programs may also be obtained from their respective sources and modified to meet the user’s specific

needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both 5 unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

10 A “recombinant nucleic acid” is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have 15 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

20 Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

25 A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

30 An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

35 The term “sample” is used in its broadest sense. A sample suspected of containing KPP, nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell,

chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or 5 synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

10 The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

15 A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

20 A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid 25 sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well 30 as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor 35 of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with

a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be 5 introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having 10 at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater 15 sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the 20 reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide 25 polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-30 1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

35 THE INVENTION

The invention is based on the discovery of new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

5 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is
10 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte 15 polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

20 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the
25 MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these 30 properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:1 is 33% identical to human MAP kinase phosphatase 6 (GenBank ID g6840994) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.3e-19, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a dual specificity phosphatase catalytic domain as determined
35 by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM

database of conserved protein family domains. (See Table 3.) Data from BLAST analysis of the DOMO database provides further corroborative evidence that SEQ ID NO:1 is a MAP kinase phosphatase.

In an alternative example, SEQ ID NO:2 is 86% identical, from residue M1 to residue 5 M2299, to Rattus norvegicus glomerular mesangial cell receptor protein-tyrosine phosphatase precursor (GenBank ID g3300096) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a protein-tyrosine phosphatase domain as determined by searching for statistically significant matches 10 in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:2 is a protein-tyrosine phosphatase.

In an alternative example, SEQ ID NO:3 is 96% identical, from residue M1 to residue Q316, to a mouse tau tubulin kinase (GenBank ID g15341198) as determined by the Basic Local Alignment 15 Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.0e-179, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, and additional BLAST analyses provide further 20 corroborative evidence that SEQ ID NO:3 is a protein kinase.

In an alternative example, SEQ ID NO:4 is 98% identical, from residue H234 to residue C1863, to human nuclear dual-specificity phosphatase (GenBank ID g3015538) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. 25 SEQ ID NO:4 also contains a DENN (AEX-3) domain and a pleckstrin homology domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analysis using the PRODOM database provides further corroborative evidence that SEQ ID NO:4 is a nuclear phosphatase.

30 In an alternative example, SEQ ID NO:5 is 76% identical, from residue M1 to residue T382, to mouse serine/threonine kinase 33 (GenBank ID g14148952) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.0e-163, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a protein kinase domain as determined by searching for statistically significant 35 matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family

domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:5 is a protein kinase.

In an alternative example, SEQ ID NO:6 is 42% identical, from residue V107 to residue L331, to a Caenorhabditis elegans protein similar to the protein phosphatases 2c family (GenBank ID g2804429) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.4e-41, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a protein phosphatase 2C domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and BLAST_DOMO analyses provide further corroborative evidence that SEQ ID NO:6 is a protein phosphatase 2C.

In an alternative example, SEQ ID NO:7 is 75% identical, from residue F46 to residue V292, to Mus musculus PFTAIRE kinase (GenBank ID g2392814) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.3e-103, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 is also 73% identical, from residue F46 to residue R296, to human serine/threonine protein kinase PFTAIRE-1 (GenBank ID g12002201), with a BLAST probability score of 9.0e-103. SEQ ID NO:7 also contains a protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:7 is a protein kinase.

In an alternative example, SEQ ID NO:8 is 58% identical, from residue K2 to residue L281, to human pyrimidine 5'-nucleotidase (GenBank ID g11245474) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.1e-90, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. The algorithms and parameters for the analysis of SEQ ID NO:1-8 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:9-16 or that

distinguish between SEQ ID NO:9-16 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation “ENST”). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation “NP”). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a “stitched” sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYY is the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAAA_gBBBBBB_1_N is a “stretched” sequence, with XXXXXX being the Incyte project identification number, gAAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, gBBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).

GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

5 In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is 10 the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

15 The invention also encompasses KPP variants. A preferred KPP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the KPP amino acid sequence, and which contains at least one functional or structural characteristic of KPP.

The invention also encompasses polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected 20 from the group consisting of SEQ ID NO:9-16, which encodes KPP. The polynucleotide sequences of SEQ ID NO:9-16, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding KPP. In 25 particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-16 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% 30 polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:9-16. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of KPP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant

of a polynucleotide sequence encoding KPP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding KPP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 5 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding KPP. Any one of the splice variants described above can 10 encode an amino acid sequence which contains at least one functional or structural characteristic of KPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be 15 produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

20 Although nucleotide sequences which encode KPP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding KPP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide 25 occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

30 The invention also encompasses production of DNA sequences which encode KPP and KPP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding KPP or any fragment thereof.

35 Also encompassed by the invention are polynucleotide sequences that are capable of

hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:9-16 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in 5 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASe (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or 10 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA 15 sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

20 The nucleic acid sequences encoding KPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)
25 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, 30 M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries 35 (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries

and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of 5 about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence 10 into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the 15 emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

20 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode KPP may be cloned in recombinant DNA molecules that direct expression of KPP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express KPP.

25 The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction 30 sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or 35

improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding KPP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.)

Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active KPP, the nucleotide sequences encoding KPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where

sequences encoding KPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the 5 vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression 10 vectors containing sequences encoding KPP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

15 A variety of expression vector/host systems may be utilized to contain and express sequences encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, 20 or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New 25 York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

35 In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding KPP can be achieved using a

multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSORT1 plasmid (Life Technologies). Ligation of sequences encoding KPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for 5 in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

10 Yeast expression systems may be used for production of KPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 15 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of KPP. Transcription of sequences encoding KPP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology 25 (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding KPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain 30 infective virus which expresses KPP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of 35 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are

constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression 5 of KPP in cell lines is preferred. For example, sequences encoding KPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance 10 to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine 15 phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; Lowy, I. et al. (1980) *Cell* 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., 20 Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate 25 luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the 30 sequence encoding KPP is inserted within a marker gene sequence, transformed cells containing sequences encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

35 In general, host cells that contain the nucleic acid sequence encoding KPP and that express

KPP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

5 Immunological methods for detecting and measuring the expression of KPP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive 10 binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

15 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding KPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding KPP, or any fragments thereof, may be cloned into a vector for 20 the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for 25 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence 30 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of 35 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation,

phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the 5 American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein 10 containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, 15 *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a 20 proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein sequence, so that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

25 In a further embodiment of the invention, synthesis of radiolabeled KPP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

30 KPP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to KPP. At least one and up to a plurality of test compounds may be screened for specific binding to KPP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

35 In one embodiment, the compound thus identified is closely related to the natural ligand of KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a

natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which KPP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for 5 these compounds involves producing appropriate cells which express KPP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP or the compound is analyzed.

10 An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a 15 labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

KPP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or 20 inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a 25 compound that modulates the activity of KPP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding KPP or their mammalian homologs may be 30 “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted 35 by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science*

244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330).

5 Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding KPP may also be manipulated in vitro in ES cells derived from 10 human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

15 Polynucleotides encoding KPP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. 20 Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

Therapeutics

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists 25 between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing KPP are brain, testes, uterine, breast and lymph node tissues and also can be found in Table 6. Therefore, KPP appears to play a role in cardiovascular diseases, immune system disorders, 30 neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP.

Examples of such disorders include, but are not limited to, a cardiovascular disease such as 35 arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial

dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis,

5 congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism,

10 pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis,

15 hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung

20 transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema,

25 episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic

30 anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other

35 extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive

neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-

5 Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other 10 neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, 15 and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR 20 syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural 25 hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, 30 lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypcholesterolemia, Tay-Sachs disease, Sandhoff's disease, 35 hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as

actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, 5 bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

In another embodiment, a vector capable of expressing KPP or a fragment or derivative 10 thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified KPP in 15 conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of KPP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those listed above.

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or 20 prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a 25 pharmaceutical agent to cells or tissues which express KPP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary 30 sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic 35 efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of KPP may be produced using methods which are generally known in the art. In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, 5 polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with KPP or with any fragment or oligopeptide thereof 10 which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

15 It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric 20 molecule may be produced.

Monoclonal antibodies to KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. 25 Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. 30 Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce KPP-specific single chain antibodies. Antibodies with related specificity, but of distinct idotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., 35 Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

5 Antibody fragments which contain specific binding sites for KPP may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D.
10 et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between KPP and its specific
15 antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an
20 association constant, K_a, which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a
25 particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10⁹ to 10¹² L/mole are preferred for use in immunoassays in which the KPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10⁶ to 10⁷ L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of KPP, preferably in active form, from the antibody
30 (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For
35 example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml,

preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

5 In another embodiment of the invention, the polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding KPP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments
10 can be designed from various locations along the coding or control regions of sequences encoding KPP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence
15 complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other
20 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475),
25 cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated
30 cell proliferation), or (iii) express a protein which affords protection against intracellular parasites

(e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the 5 case where a genetic deficiency in KPP expression or regulation causes disease, the expression of KPP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in KPP are treated by constructing mammalian expression vectors encoding KPP and introducing these 10 vectors by mechanical means into KPP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 15 9:445-450).

Expression vectors that may be effective for the expression of KPP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). KPP may 20 be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid 25 (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID 30 TRANSECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of 35 these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a

tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. 5 Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. 10 For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of 15 ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding KPP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During 20 alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding 25 RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the 30 introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions 35 -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly,

inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

5 Ribozyomes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme 10 molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding KPP.

15 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for 20 secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

25 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA 30 sequences encoding KPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, 35 cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible 35 modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous 40 endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding KPP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, 5 transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased KPP expression or activity, a compound which specifically inhibits expression of the polynucleotide 10 encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding KPP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method 15 commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a 20 polynucleotide encoding KPP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding KPP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence 25 of the polynucleotide encoding KPP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide 30 can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a 35 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide

sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells 5 taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) *Nat. Biotechnol.* 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of 10 such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. 15 Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, 20 intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of 25 fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

30 Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of 35 macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of

the macromolecule. Alternatively, KPP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

5 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

10 A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose 15 lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity.

20 The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the 25 severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μ g to 100,000 μ g, up to a total dose of 30 about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding KPP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding KPP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:9-16 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

Means for producing specific hybridization probes for DNAs encoding KPP include the

cloning of polynucleotide sequences encoding KPP or KPP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, 5 for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, 10 Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular 15 calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, 20 pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary 25 eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder 30 such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia 35 with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,

glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, 5 amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal 10 familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral 15 nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial 20 frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary 25 thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, 30 myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid 35 disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine

palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, 5 lipomas, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, 10 hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, 15 bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. The polynucleotide sequences encoding KPP may be used in Southern or 20 northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding KPP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding KPP may be labeled by standard methods and added to a fluid or tissue sample 25 from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding KPP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate 30 the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of KPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a 35 fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification.

Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to 5 establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several 10 days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals 15 to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide 20 encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences 25 encoding KPP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding KPP are used to amplify 30 DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high- 35 throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis

methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of KPP include radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used

to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

5 In another embodiment, KPP, fragments of KPP, or antibodies specific for KPP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of 10 gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of 15 transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, 20 biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental 25 compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share 30 those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for 35 comparison of expression data after treatment with different compounds. While the assignment of

gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at 5 <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of 10 the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present 15 invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by 20 separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by 25 staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the 30 spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

35 A proteomic profile may also be generated using antibodies specific for KPP to quantify the

levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are

well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding KPP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either 5 coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), 10 yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop 15 genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man 20 (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. 25 Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, 30 any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or 35 oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug

screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds 5 having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed. Bound KPP is then detected by methods well known in the art. Purified KPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, 10 non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding KPP specifically compete with a test compound for binding KPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

15 In additional embodiments, the nucleotide sequences which encode KPP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding 20 description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/263,083, U.S. Ser. No. 60/271,205, U.S. Ser. No. 60/271,117, U.S. Ser. 25 No. 60/276,859, U.S. Ser. No. 60/278,504, U.S. Ser. No. 60/278,522, U.S. Ser. No. 60/280,510 and U.S. Ser. No. 60/280,266 are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

30 Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with 35 chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and

ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, 5 Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP 10 vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-15 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), 20 PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

25 Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 30 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 35 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically

using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.

5 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as 10 the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading 15 frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and 20 programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, 25 Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to 30 produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length 35 polynucleotide sequences were translated to derive the corresponding full length polypeptide

sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov 5 model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also 10 calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate 15 references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide 20 and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:9-16. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases were initially identified by running the Genscan gene 25 identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of 30 Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases were also identified by homology to Incyte cDNA sequences that had been annotated as 35 kinases and phosphatases. These selected Genscan-predicted sequences were then compared by

BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing 5 evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted 10 coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example 15 III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on 20 more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear 25 along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan 30 were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public 35 databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases

using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions 5 may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

10 **VI. Chromosomal Mapping of KPP Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:9-16 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:9-16 were assembled into clusters of contiguous and overlapping sequences using 15 assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

20 Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM 25 distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

30 **VII. Analysis of Polynucleotide Expression**

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

35 Analogous computer techniques applying BLAST were used to search for identical or related

molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

5

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \min\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding KPP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding KPP. cDNA sequences and cDNA library/tissue

information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of KPP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One 5 primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would 10 result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction 15 mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C. In the 20 alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 57 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN 25 quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the 30 sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) 35 agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones

were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 5 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was 10 quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

15 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were 20 identified in SEQ ID NO:9-16 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment 25 errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated 30 algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in 35 four different human populations. The Caucasian population comprised 92 individuals (46 male, 46

female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:9-16 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g.,

Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be 5 selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a 10 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

15 Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse 20 transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified 25 using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 35 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia

Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR

5 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

10 Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

15 Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

20 Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash 25 buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

30 Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

35 In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.

Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is 5 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that 10 location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

15 The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and 20 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The 25 software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Polynucleotides

Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring KPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with 30 smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of KPP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the KPP-encoding transcript.

35 **XIII. Expression of KPP**

Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the

5 T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG).

Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as

10 baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional 15 genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-

20 kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity 25 purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX and XXI where applicable.

XIV. Functional Assays

30 KPP function is assessed by expressing the sequences encoding KPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into 35 a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome

formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP); 5 Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; 10 changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are 15 discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of KPP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding KPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using 20 magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of KPP Specific Antibodies

25 KPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

30 Alternatively, the KPP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

35 Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to

increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-
5 rabbit IgG.

XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

Naturally occurring or recombinant KPP is substantially purified by immunoaffinity chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP antibody to an activated chromatographic resin, such as CNBr-activated
10 SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing KPP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/KPP binding
15 (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and KPP is collected.

XVII. Identification of Molecules Which Interact with KPP

KPP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules
20 previously arrayed in the wells of a multi-well plate are incubated with the labeled KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of KPP are used to calculate values for the number, affinity, and association of KPP with the candidate molecules.

Alternatively, molecules interacting with KPP are analyzed using the yeast two-hybrid
25 system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S.
30 Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by KPP in the presence of $[\gamma-^{32}\text{P}]$ ATP. KPP is incubated with the protein substrate,
35 ^{32}P -ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P -ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The

amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P -ATP. The reservoir of the centrifuged unit containing the ^{32}P -peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and p34^{cdc2}kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in vitro in an assay containing KPP, 50 μl of kinase buffer, 1 μg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 μg ATP, and 0.5 μCi [γ - ^{32}P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ - ^{32}P]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a 12% SDS polyacrylamide gel followed by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the presence of the inactive kinase, K38A. The amount of incorporated ^{32}P is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of ^{32}P from [γ - ^{32}P]ATP into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and ^{32}P -labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is proportional to the activity of KPP.

In yet another alternative, other assays for KPP include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant

proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β -mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively, acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 μ l of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μ l of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of KPP in the assay.

In the alternative, KPP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30 μ l containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β -mercaptoethanol and 10 μ M substrate, 32 P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μ l of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM Na₄P₂O₇, and 2 mM NaH₂PO₄, then centrifuged at 12,000 \times g for 5 min. Acid-soluble 32 Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of 125 I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated 32 P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors and Activators

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVIII. KPP activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules

which enhance KPP activity.

Agonists or antagonists of KPP activation or inhibition may be tested using assays described in section XVIII. Agonists cause an increase in KPP activity and antagonists cause a decrease in KPP activity.

5 XXI. Identification of KPP Substrates

A KPP “substrate-trapping” assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a 10 clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in Escherichia coli, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in E. coli and purified by 15 chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 μ g of CsCl-purified DNA per 10-cm dish of cells or 8 μ g per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are 20 lysed in 50 mM Tris-HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μ g/ml leupeptin/5 μ g/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 μ g of mAb or 10 μ l of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or 25 further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the 30 invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
55074884	1	55074884CD1	9	55074884CB1
7480588	2	7480588CD1	10	7480588CB1
7482931	3	7482931CD1	11	7482931CB1
2080788	4	2080788CD1	12	2080788CB1
71918969	5	71918969CD1	13	71918969CB1
8187571	6	8187571CD1	14	8187571CB1
7494145	7	7494145CD1	15	7494145CB1
5807954	8	5807954CD1	16	5807954CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	55074884CD1	g6840994	1.3E-19	[Homo sapiens] MAP kinase phosphatase 6 Marti, F. et al. (2001) J Immunol 166:197-206
2	7480588CD1	g3300096	0	[Rattus norvegicus] glomerular mesangial cell receptor protein-tyrosine phosphatase precursor Wright, M.B. et al. (1998) J. Biol. Chem. 273:23929-23937
3	7482931CD1	g15341198	1.0E-179	[Mus musculus] Tau-tubulin kinase Tomizawa, K., et al. (2001) FEBS letters 492:221-227
4	2080788CD1	g3015538	0	[Homo sapiens] nuclear dual-specificity phosphatase Cui, X. et al. (1998) Nature Genet. 18:331-337
5	71918969CD1	g14148952	1.0E-163	[Mus musculus] Serine/Threonine kinase 33 Mujica, A.O., et al. (2001) Gene 280:175-181
6	8187571CD1	g2804479	5.4E-41	[Caenorhabditis elegans] similar to the protein phosphatase 2c family
7	7494145CD1	g2392814	1.3E-103	[Mus musculus] PFTAIRE kinase Lazzaro, M.A. et al. (1997) J. Neurochem. 69:348-364.
8	5807954CD1	g11245474	1.1E-90	[Homo sapiens] pyrimidine 5'-nucleotidase Amici, A. et al. (2000) Blood 96:1596-1598

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs		Analytical Methods and Databases
1	55074884CD1	176	S30 S135	N50	Signal cleavage: M1-A36		SPSCAN
					Signal Peptide: M1-G35		HMMER
					Dual specificity phosphatase, catalytic domain: P18-A156		HMMER_PFAM
					VHL-TYPE DUAL SPECIFICITY PHOSPHATASE		BLAST_DOMO
					DM03823		
					P28562 I69-314; V20-A159		
					I38890 29-320; V20-L157		
					A56115 51-336; V20-A159		
					Q02256 I-174; G97-E155		
2	7480588CD1	2299	S104 S182 S186	N155 N162 N311	Signal Peptide: M1-T17, M1-V19		HMMER
			S200 S298 S332	N349 N384 N575			
			S350 S490 S557	N713 N731 N765			
			S567 S649 S661	N770 N809 N895			
			S668 S687 S715	N903 N959 N988			
			S733 S751 S761	N996 N1008			
			S772 S862 S898	N1038 N1059			
			S912 S927 S1277	N1071 N1161			
			S1312 S1338	N1194 N1248			
			S1381 S1385	N1253 N1292			
			S1411 S1530	N1380 N1399			
			S1558 S1590	N1524 N1584			
			S1628 S1836	N1594 N1618			
			S1900 S1941	N1647 N1694			
			S1975 S1979	N1719 N1802			
			S1984 S2018	N2059 N2162			
			S2237 S2261	N2273			

Table 3

SEQ	Incyte ID NO: Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2 (cont.)			T17 T28 T120 T208 T249, T313 T328 T365 T389 T413 T511 T607 T614 T623 T633 T660 T678 T757 T767 T824 T834 T849 T870 T907 T998 T1013 T1048 T1095 T1129 T1145 T1158 T1223 T1334 T1424 T1635 T1759 T1770 T1884 T1894 T1959 T2020 T2082 T2143 T2157		Protein-tyrosine phosphatase: N2027-D2258	HMMER_PFAM
		Y710 Y1940			Fibronectin type III domain: P51-S138, P665-V746, P854-R936, P759-S842, P1340-S1418, P1150-S1231, L1243-S1328, P1430-S1526, P948-S1040, P565-S651, P1538-S1620, P300-S383, P1641-V1734, P1053-S1134, P150-G286, P394-T554	HMMER_PFAM
					Predicted transmembrane segments: I1604-A1622, R1902-II1930	TMAP
					Tyrosine specific protein phosphatases proteins BL00383: K2030-V2044, S2055-I2063, Q2160-P2172, V2198-G2208, R2236-F2251	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2 (cont.)					Receptor tyrosine kinase class V proteins BL00790: G66-V91, N115-T145	BLIMPS_BLOCKS
					Tyrosine specific protein phosphatases signature and profiles tyr_phosphatase.prf: L2178-R2236	PROFILESCAN
					Fibronectin type III repeat signature PR00014: T1165-P1174, Y1402-D1416	BLIMPS_PRINTS
					Protein tyrosine phosphatase signature PR00700: D2056-I2063, F2072-E2092, M2156-E2173, P2195-L2213, V2226-C2241, M2242-L2252	BLIMPS_PRINTS
					GLOMERULAR MESANGIAL CELL RECEPTOR PROTEINTYROSINE PHOSPHATASE PRECURSOR PD178547: C1913-W2026 PD174021: W1528-K1642 PD184208: T145-D205	BLAST_PRODOM
					HYDROLASE PHOSPHATASE PROTEIN TYROSINE PRECURSOR PD000167: N2027-N2221	BLAST_PRODOM
					PROTEIN-TYROSINE-PHOSPHATASE DM00089	BLAST_DOMO
					A57064 896-1170: H1991-I2256 S60613 924-1198: H1991-I2256 I49372 113-387: H1991-I2256 I49374 1-269: L1995-I2256	
					ATP/GTP-binding site motif A (P-loop) AI621- S1628	MOTIFS

Table 3

SEQ	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2 (cont.)					Tyrosine specific protein phosphatases active site: V2198-F2210	MOTIFS
3	7482931CD1	478	S173 S237 S263 S278 S283 S353 T118 T155 T189 T269 T300 T402 T431 T456 T463 Y471		Eukaryotic protein kinase domain: W21-S256	HMMER_PFFAM
					SIMILAR TO CASEIN KINASES PD115501: F213-E298, L11-T114	BLAST_PRODOM
					PROTEIN KINASE DOMAIN DM00004 P42169 74-330: I27-E240 DM00004 P48730 11-265: R25-Y273 DM00004 P355506 19-273: V23-Y273 DM00004 P40235 13-267: R25-Y273	BLAST_DOMO
					Protein kinases ATP-binding region signature I27-K50	MOTIFS

Table 3

SEQ	IncYTE ID NO: Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	2080788CD1	1867	S75 S137 S182 S359 S430 S559 S637 S659 S716 S722 S773 S803 S812 S873 S968 S1047 S1094 S1113 S1120 S1174 S1199 S1232 S1237 S1259 S1280 S1284 S1362 S1387 S1465 S1517 S1527 S1552 S1561 S1630 T84 T95 T96 T102 T313 T538 T636 T933 T971 T1069 T1599 T1747 T1827 T1838 T1839 Y1769	N1585 N1761	DENN (AEX-3) domain: L171-G310	HMMER_Pfam

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4 (cont.)					HYDROLASE PROTEIN MYOTUBULARIN DISEASE MUTATION F53A2.8 PROTEIN TYROSINE PHOSPHATASE C19A8.03 CPA2NNF1 PD014611: P1379-H1470, D1129-G1266, G1474-Y1556, G927-F938	BLAST_PRODOME
					PROTEIN REGULATOR OF PRESYNAPTIC ACTIVITY SERINE PROTEASE INHIBITOR RAB3 GDP/GTP PD008900: P177-D340, R3-H115, D364-R408	BLAST_PRODOME
5	71918969CD1	514	S4, S10, S20, S29, S71, S80, S87, S169, S211, S249, S274, S296, S326, S349, S388, S407, S420, S441, S473, T11, T70, T134, T184, T371, T412, T440, T450, T459, T510	N95, N213, N411, N438, Eukaryotic protein kinase domain: Y116-L381	MOTIFS	HMMER-PFAM
					Protein kinases signatures and profile (protein_kinase_tyr.prf): E214-A269	PROFILESCAN
					Tyrosine kinase catalytic domain signature: PR00109: M192-D205, Y228-V246, C304-S326, T184-R206	BLIMPS-PRINTS
					PROTEIN KINASE DOMAIN: DM00004 S57347 21-266: F118-T371	BLAST-DOMO

Table 3

SEQ	Incyte ID NO: Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5 (cont.)					Protein kinases ATP-binding region signature: L122-K145	MOTIFS
6	8187571CD1	338	S55 S220 S281 S320 T134 T155 T246 T268 T305 Y248	N178	Serine/Threonine protein kinases active-site signature: I234-V246 Protein phosphatase 2C: S63-R128, V212-A282	MOTIFS
					Transmembrane domain: V89-I117 N terminus is non-cytosolic.	TMAP
					Protein phosphatase 2C proteins BL01032: L64-A81, N109-E148, R205-D218, D253-D265, S320-I329	BLIMPS_BLOCKS
					PROTEIN PHOSPHATASE 2C DM00377P4959611-295: S55-G143, R205-S281 DM00377Q0917311-296: V66-L141, E188-A282 DM00377S6246211-297: V66-L141, E188-A282	BLAST_DOMO
7	7494145CD1	321	S19 S24 S228 Y63		Eukaryotic protein kinase domain: Y52-P275	HMMER_PFAM
					Transmembrane domain: D231-L248; N-terminus is cytosolic	TMAP
					Protein kinases signatures and profile (protein_kinase_tyr.prf): N149-A197	PROFILESCAN
					KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP-BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN: PD00268: Q203-P281	BLAST-PRODOM
					PROTEIN KINASE DOMAIN: DM00004Q00536166-436: L53-L305	BLAST-DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP-BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN: PD002668; Q203-P281	BLAST-PRODOM
7 (cont.)					PROTEIN KINASE DOMAIN: DM000041Q005361L166-436; L53-L305	BLAST-DOMO
					Protein kinases ATP-binding region signature: L58-K81	MOTIFS
8	5807954CD1	292	S32 S200 S223 T86 T228	N198	Serine/Threonine protein kinases active-site signature: V169-I181	MOTIFS
					Transmembrane domain: N135-K163 N-terminus is cytosolic	TMAP

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
9/55074884CB1/ 1605	1-361, 1-830, 4-517, 188-830, 346-731, 437-731, 440-493, 440-495, 440-516, 440-517, 440-528, 440-731, 440-1205, 440-1285, 443-731, 450-731, 480-731, 520-774, 540-774, 541-731, 544-731, 596-731, 616-731, 645-731, 651-731, 672-731, 683-731, 777-1605, 1026-1191, 1052-1191, 1102-1377, 1337-1440
10/7480588CB1/ 7225	1-148, 1-460, 159-377, 376-894, 646-1584, 1351-2448, 1517-1882, 1700-2082, 1883-1955, 1883-2556, 1883-3153, 2242-2678, 2272-2680, 2719-3351, 2986-3351, 3154-3444, 3171-3477, 3171-3513, 3171-3522, 3171-3551, 3171-3604, 3171-3620, 3171-3646, 3171-3680, 3171-3689, 3179-3689, 3202-3689, 3220-3682, 3242-3689, 3244-3689, 3246-3691, 3446-4285, 3459-4284, 4234-4915, 4234-4925, 4235-4920, 4237-4925, 4245-4925, 4254-4925, 4271-4925, 4288-5687, 4291-4925, 4297-4925, 4303-4925, 4307-4925, 4324-4925, 4330-4909, 4330-4678, 4330-4685, 4382-4677, 4382-4681, 4382-4682, 4382-4684, 4382-4685, 4384-4684, 4405-4683, 4435-4684, 4467-4680, 5110-5794, 5110-5852, 5110-5854, 5110-5862, 5110-5908, 5110-5916, 5111-5854, 5230-6114, 5444-6104, 5686-6861, 6047-6781, 6047-6784, 6047-6834, 6047-6840, 6047-6854, 6047-6863, 6047-6885, 6047-6891, 6470-7225, 6495-7225, 6507-7225, 6510-7225, 6541-7225, 6545-7225, 6567-7225, 6582-7225, 6738-6864, 6861-6883
11/7482931CB1/ 1691	1-243, 1-576, 32-235, 93-230, 95-378, 186-377, 307-703, 307-763, 307-803, 307-836, 307-845, 307-1049, 307-1056, 307-1063, 307-1076, 307-1078, 307-1080, 307-1112, 307-1127, 308-845, 311-372, 377-787, 377-797, 427-797, 465-1132, 577-1082, 757-1630, 816-1618, 822-1608, 826-1624, 878-1567, 887-1314, 907-1614, 933-1614, 935-1167, 940-1608, 982-1543, 1020-1691, 1021-1691, 1038-1691, 1052-1604, 1053-1431, 1053-1614, 1053-1642, 1053-1667, 1053-1680, 1053-1691, 1054-1691, 1055-1680, 1055-1691, 1063-1691, 1072-1691, 1163-1691
12/2080788CB1/ 6146	1-540, 4-335, 4-580, 11-216, 11-464, 19-359, 19-647, 240-412, 265-595, 271-719, 343-931, 446-1064, 522-1140, 539-821, 617-1206, 661-1241, 664-977, 696-1261, 696-1293, 700-1230, 756-1259, 756-1364, 801-1374, 814-1393, 848-1376, 953-1448, 956-1526, 974-1304, 1005-1465, 1005-1655, 1192-1838, 1206-1743, 1234-1784, 1244-1439, 1258-1819, 1259-1534, 1263-1524, 1271-1543, 1302-1734, 1307-1906, 1319-1672, 1362-1865, 1363-1784, 1368-1749, 1377-1610, 1378-1589, 1378-1640, 1378-1643, 1388-2021, 1415-1689, 1415-1756, 1415-1816, 1415-1986, 1429-1972, 1466-1973, 1475-1756, 1501-1769, 1525-1926, 1527-2112, 1539-1844, 1559-2049, 1584-1782, 1584-2191, 1584-2213, 1586-1828, 1586-2154, 1590-2153, 1603-2222, 1616-2168, 1643-2102, 1671-2212, 1692-2313, 1715-2249, 1734-2404, 1743-2341, 1754-2015, 1777-2361, 1779-2378, 1800-2352, 1800-2422, 1806-2253, 1806-2271, 1812-2343, 1836-2130, 1837-2140, 1840-2110, 1845-2434, 1846-2188, 1848-2188, 1848-2387, 1863-2420, 1878-2304, 1883-2411, 1889-2475, 1890-2426, 1892-2508, 1901-2401, 1906-2401, 1909-2415, 1919-2169, 1919-2436, 1921-2333, 1922-2271, 1945-2507, 1951-2454, 1972-2417, 1987-2515, 2028-2611, 2036-2562, 2036-2641, 2041-2292, 2041-2337, 2043-2658, 2045-2620, 2068-2509, 2100-

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
12 (cont.)	2699, 2134-2674, 2149-2396, 2149-2397, 2149-2416, 2149-2425, 2149-2427, 2149-2428, 2149-2431, 2158-2360, 2170-2734, 2180- 2769, 2188-2433, 2188-2550, 2188-2648, 2199-2532, 2206-2847, 2208-2847, 2213-2847, 2216-2790, 2218-2764, 2221-2693, 2226- 2847, 2229-2842, 2229-2847, 2230-2602, 2231-2847, 2235-2842, 2235-2847, 2242-2847, 2247-2790, 2255-2842, 2258-2847, 2259- 2847, 2266-2847, 2290-2847, 2295-2780, 2296-2846, 2308-2847, 2309-2842, 2310-2605, 2312-2847, 2314-2596, 2321-2842, 2322- 2586, 2337-2847, 2339-2616, 2348-2838, 2356-2816, 2357-2847, 2360-2829, 2364-2847, 2365-2824, 2365-2847, 2366-2847, 2371- 2847, 2375-2847, 2380-2838, 2383-2847, 2386-2754, 2400-2847, 2406-2838, 2413-2829, 2413-2841, 2417-2829, 2417-2847, 2421- 2847, 2424-2756, 2425-2500, 2431-2658, 2441-2780, 2443-2847, 2452-2846, 2476-2847, 2478-2847, 2479-3129, 2482-2785, 2487- 2784, 2496-2735, 2496-2847, 2507-2846, 2523-2744, 2523-2847, 2559-2842, 2540-2847, 2559-3031, 2569-2847, 2572-2847, 2575- 2847, 2577-2847, 2601-2811, 2637-2847, 2719-3332, 2751-2828, 2801-3359, 2840-3228, 2840-3266, 2840-3288, 2840-3293, 2840- 3303, 2840-3443, 2840-3452, 2843-3221, 2908-3491, 2928-3321, 2936-3257, 2981-3245, 3333-3753, 3354-3585, 3398-3932, 3446- 3942, 3447-3714, 3617-4204, 3637-3753, 3665-3887, 3748-3772, 3905-4425, 4011-4321, 4045-4664, 4173-4445, 4186-4463, 4197- 4459, 4244-4523, 4250-4489, 4250-4794, 4255-4510, 4296-4536, 4296-4814, 4302-4514, 4303-4596, 4304-4773, 4385-4655, 4398- 4632, 4493-4768, 4519-4771, 4526-4784, 4526-5055, 4538-4940, 4542-4689, 4542-4924, 4550-4784, 4589-5001, 4594-4855, 4640- 4899, 4669-4958, 4689-4874, 4731-5023, 4737-4982, 4757-5017, 4764-5064, 4777-5026, 4777-5196, 4784-5043, 4788-5087, 4793- 5051, 4793-5297, 4800-5254, 4819-5107, 4832-5140, 4846-5072, 4901-5155, 4907-5184, 4910-5177, 4960-5351, 5006-5236, 5020- 5245, 5022-5314, 5022-5632, 5040-5324, 5055-5256, 5123-5372, 5135-5405, 5153-5394, 5183-5481, 5213-5456, 5214-5456, 5216- 5500, 5226-5424, 5282-5555, 5311-5569, 5320-5559, 5322-5524, 5322-5798, 5368-5623, 5380-5698, 5440-5914, 5452-5654, 5463- 5709, 5463-6000, 5464-5722, 5468-5722, 5468-5815, 5484-5784, 5502-5736, 5525-5722, 5525-5786, 5536-5892, 5539-6109, 5543- 5800, 5544-6102, 5554-6085, 5588-6077, 5601-5813, 5601-5910, 5604-5739, 5604-5864, 5604-5871, 5604-6115, 5605-5876, 5619- 6103, 5637-5889, 5638-6146, 5667-6104, 5669-5853, 5673-5918, 5740-5970, 5760-6048, 5790-6128, 5909-6092, 5933-6130, 5939- 6127, 5946-6124, 5950-6124, 5981-6131

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
13/71918969CB1/ 2362	1-341, 1-450, 1-488, 1-546, 1-590, 1-753, 25-325, 25-443, 30-341, 66-338, 249-638, 249-958, 251-597, 457-951, 457-969, 457-1091, 520-785, 523-1161, 580-1032, 606-1177, 625-1147, 630-883, 631-1368, 634-1297, 755-1422, 762-1323, 816-1493, 818-1529, 822-1520, 829-1113, 829-1339, 842-1520, 847-1120, 865-1395, 915-1584, 920-1638, 936-1575, 951-1623, 995-1622, 1003-1594, 1021-1735, 1030-1620, 1082-1820, 1086-1769, 1111-1785, 1135-1309, 1145-1309, 1169-1309, 1169-1769, 1192-1620, 1198-1897, 1205-1899, 1209-1309, 1218-1785, 1220-1920, 1222-1460, 1222-1505, 1223-1309, 1224-1309, 1235-1307, 1235-1754, 1243-1831, 1247-1309, 1248-1882, 1250-1309, 1253-1880, 1276-1961, 1277-1908, 1291-1866, 1301-1703, 1311-1929, 1358-1948, 1386-1926, 1454-1969, 1455-1698, 1468-1871, 1477-2276, 1481-1677, 1482-2128, 1498-2010, 1504-2130, 1504-2152, 1511-2169, 1511-2245, 1544-2319, 1557-2310, 1566-2134, 1580-2240, 1581-2325, 1590-2152, 1597-2132, 1602-2325, 1620-2266, 1624-2171, 1632-2277, 1643-2325, 1650-1903, 1660-2151, 1666-2054, 1685-2325, 1686-1906, 1686-1920, 1686-2163, 1686-2206, 1699-1925, 1718-2174, 1722-2171, 1748-2020, 1749-2170, 1756-2338, 1761-2325, 1783-2042, 1804-2168, 1805-2166, 1812-2212, 1827-2362, 1828-2173, 1890-2325, 1900-2325, 1901-2077, 1903-2325, 1920-2166
14/8187571CB1/ 1535	1-617, 142-828, 175-330, 178-348, 181-459, 239-814, 258-487, 270-795, 311-855, 400-640, 400-983, 557-812, 711-938, 715-988, 783-1079, 803-1371, 810-1509, 834-1431, 886-1528, 1022-1519, 1089-1533, 1124-1535, 1131-1527, 1133-1379, 1145-1373, 1182-1531, 1188-1531, 1215-1336, 1215-1518, 1215-1535, 1216-1531, 1242-1533, 1254-1528, 1259-1325, 1277-1532, 1294-1526, 1330-1531, 1331-1535, 1343-1531
15/7494145CB1/ 1376	1-469, 149-469, 153-207, 161-213, 189-521, 268-417, 268-778, 268-812, 268-951, 268-982, 556-1142, 556-1145, 556-1147, 556-1149, 568-1149, 842-1147, 864-1154, 904-1376
16/5807954CB1/ 1482	1-41, 1-298, 14-485, 34-381, 69-264, 144-642, 191-891, 196-438, 238-482, 259-470, 259-473, 271-575, 283-635, 293-581, 293-961, 359-570, 359-639, 378-639, 382-1008, 438-1020, 483-762, 562-891, 635-879, 640-895, 653-1172, 654-941, 654-1003, 691-938, 886-1117, 906-1155, 909-1156, 918-1146, 929-1191, 932-1122, 932-1178, 932-1186, 932-1238, 934-1173, 945-1136, 995-1429, 1007-1203, 1007-1272, 1009-1227, 1009-1236, 1009-1381, 1009-1408, 1009-1447, 1138-1477, 1170-1333, 1237-1482, 1284-1477

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
9	55074884CB1	HEARNON03
10	7480588CB1	LUNGFER04
11	7482931CB1	BRAQNOT01
12	2080788CB1	BRAUNOR01
13	71918969CB1	TESTNOT03
14	8187571CB1	UTRSNOT01
15	7494145CB1	BRSTTMC01
16	5807954CB1	LNODNOT03

Table 6

Library	Vector	Library Description
BRAQNOT01	pINCY	Library was constructed using RNA isolated from midbrain tissue removed from a 35-year-old Caucasian male. No neuropathology was found. Patient history included dilated cardiomyopathy, congestive heart failure, and an enlarged spleen and liver.
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloid goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRSTTMC01	pINCY	This large size-fractionated library was constructed using pooled cDNA from four donors. cDNA was generated using mRNA isolated from diseased breast tissue removed from a 40-year-old Caucasian female (donor A) during a bilateral reduction mammoplasty; from breast tissue removed from a 46-year-old Caucasian female (donor B) during unilateral extended simple mastectomy with breast reconstruction; from breast tissue removed from a 56-year-old Caucasian female (donor C) during unilateral extended simple mastectomy with open breast biopsy; and from breast tissue removed from a 57-year-old Caucasian female (donor D) during a unilateral extended simple mastectomy. Pathology indicated bilateral mild fibrocystic and proliferative changes (A); deep fascia was negative for tumor (B); non-proliferative fibrocystic change (C); and benign fat replaced breast parenchyma (D). Pathology for the matched tumor tissue (B) indicated invasive grade 3 adenocarcinoma, ductal type, with apocrine features. Pathology for the matched tumor tissue (C) indicated invasive grade 3 ductal adenocarcinoma. Pathology for the matched tumor tissue (D) indicated residual microscopic infiltrating grade 3

Table 6

Library	Vector	Library Description
BRSTTMC01 (cont.)		ductal adenocarcinoma and extensive grade 2 intraductal carcinoma. Patient history included breast hypertrophy and pure hypercholesterolemia (A); breast cancer (B); chronic airway obstruction and emphysema (C); and benign hypertension, hyperlipidemia, cardiac dysrhythmia, a benign colon neoplasm, a solitary breast cyst, and a breast neoplasm of uncertain behavior (D). Previous surgeries included open breast biopsy (B). Donor B's medications included Cytosan and Adriamycin.
HEARNON03	pINCY	This normalized heart tissue library was constructed from 8.4 million independent clones from a heart tissue library. Starting RNA was made from heart tissue removed from a 44-year-old Caucasian male, who died from intracranial hemorrhage. Serology was positive for anti-CMV (cytomegalovirus). Patient history included back and neck pain, hypertension, pneumonia, sinus infection, alcohol use, and daily pipe tobacco use (x3 years). Patient medications included Procardia. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LNODNOT03	pINCY	Library was constructed using RNA isolated from lymph node tissue obtained from a 67-year-old Caucasian male during a segmental lung resection and bronchoscopy. On microscopic exam, this tissue was found to be extensively necrotic with 10% viable tumor. Pathology for the associated tumor tissue indicated invasive grade 3-4 squamous cell carcinoma. Patient history included hemangioma. Family history included atherosclerotic coronary artery disease, benign hypertension, congestive heart failure, atherosclerotic coronary artery disease.
LUNGFER04	PCDNA2.1	This random primed library was constructed using RNA isolated from lung tissue removed from a Caucasian male fetus who died from fetal demise.
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
UTRSNOT01	PSPORT1	Library was constructed using RNA isolated from the uterine tissue of a 59-year-old female who died of a myocardial infarction. Patient history included cardiomyopathy, coronary artery disease, previous myocardial infarctions, hypercholesterolemia, hypotension, and arthritis.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABIPARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tblastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	<i>ESTs</i> : Probability value= 1.0E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value= 1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 266:88-105; and Attwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:320-322; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nutshell, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197, and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and
 - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

5 b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

5

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

10 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16,
15 b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16,
c) a polynucleotide complementary to a polynucleotide of a),
d) a polynucleotide complementary to a polynucleotide of b), and
e) an RNA equivalent of a)-d).

20 13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

25 14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

25 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

30 15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

35 a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and

- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

5 17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

10 19. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition of claim 17.

15 20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20 21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

25 22. A method for treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

30 24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

35 25. A method for treating a disease or condition associated with overexpression of functional

KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

10 27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 15 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

20 28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 25 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

30

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target

polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

5 c) quantifying the amount of hybridization complex, and

 d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of KPP in a
10 biological sample, the method comprising:

15 a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

20 a) a chimeric antibody,
 b) a single chain antibody,
 c) a Fab fragment,
 d) a F(ab')₂ fragment, or
 e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

25 33. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

30 34. A composition of claim 32, wherein the antibody is labeled.

35 35. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

10

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

15

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

20

25

40. A monoclonal antibody produced by a method of claim 39.

30

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

35

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant

immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in a sample, the method comprising:

- 5 a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in the sample.

10

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 from a sample, the method comprising:

- 15 a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim

20 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- 25 a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

30 48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

35

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

10 52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

15 54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

20 55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

25

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

30

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

35

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

5 64. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:9.

65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:10.

10 66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.

67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

15 68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

25 71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

<110> INCYTE GENOMICS, INC.
LEE, Ernestine A.
WALIA, Narinder K.
BAUGHN, Mariah R.
ISON, Craig H.
GURURAJAN, Rajagopal
ARVIZU, Chandra
YAO, Monique G.
HILLMAN, Jennifer L.
TANG, Y. Tom
YUE, Henry
TRAN, Bao
DING, Li
LU, Dyung Aina
LAL, Preeti G.
WARREN, Bridget A.

<120> KINASES AND PHOSPHATASES

<130> PI-0344 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/263,083; 60/271,205; 60/271,117; 60/276,859; 60/278,504;
60/278,522; 60/280,510; 60/280,266
<151> 2001-01-18; 2001-02-23; 2001-02-23; 2001-03-16; 2001-03-23;
2001-03-23; 2001-03-29; 2001-03-29

<160> 16

<170> PERL Program

<210> 1
<211> 176
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55074884CD1

<400> 1
Met Gly Pro Ala Glu Ala Gly Arg Arg Gly Ala Ala Ser Pro Val
1 5 10 15
Pro Pro Pro Leu Val Arg Val Ala Pro Ser Leu Phe Leu Gly Ser
20 25 30
Ala Arg Ala Ala Gly Ala Glu Glu Gln Leu Ala Arg Ala Gly Val
35 40 45
Thr Leu Cys Val Asn Val Ser Arg Gln Gln Pro Gly Pro Arg Ala
50 55 60
Pro Gly Val Ala Glu Leu Arg Val Pro Val Phe Asp Asp Pro Ala
65 70 75
Glu Asp Leu Leu Ala His Leu Glu Pro Thr Cys Ala Ala Met Glu
80 85 90
Ala Ala Val Arg Ala Gly Gly Ala Cys Leu Val Tyr Cys Lys Asn
95 100 105
Gly Arg Ser Arg Ser Ala Ala Val Cys Thr Ala Tyr Leu Met Arg
110 115 120
His Arg Gly Leu Ser Leu Ala Lys Ala Phe Gln Met Val Lys Ser
125 130 135
Ala Arg Pro Val Ala Glu Pro Asn Pro Gly Phe Trp Ser Gln Leu
140 145 150
Gln Lys Tyr Glu Glu Ala Leu Gln Ala Gln Ser Cys Leu Gln Gly

	155	160	165
Glu Pro Pro Ala	Leu Gly Leu Gly Pro	Glu Ala	
	170	175	
<210> 2			
<211> 2299			
<212> PRT			
<213> Homo sapiens			
<220>			
<221> misc_feature			
<223> Incyte ID No: 7480588CD1			
<400> 2			
Met Asp Phe Leu Ile Ile Phe Leu Leu Leu Phe Ile Gly Thr Ser			
1 5 10 15			
Glu Thr Gln Val Asp Val Ser Asn Val Val Pro Gly Thr Arg Tyr			
20 25 30 35			
Asp Ile Thr Ile Ser Ser Ile Ser Thr Thr Tyr Thr Ser Pro Val			
35 40 45 50			
Thr Arg Ile Gly Glu Pro Gly Pro Pro Val Phe Leu Ala Gly Glu			
50 55 60 65			
Arg Val Gly Ser Ala Gly Ile Leu Leu Ser Trp Asn Thr Pro Pro			
65 70 75 80			
Asn Pro Asn Gly Arg Ile Ile Ser Tyr Ile Val Lys Tyr Lys Glu			
80 85 90 95			
Val Cys Pro Trp Met Gln Thr Val Tyr Thr Gln Val Arg Ser Lys			
95 100 105 110			
Pro Asp Ser Leu Glu Val Leu Leu Thr Asn Leu Asn Pro Gly Thr			
110 115 120 125			
Thr Tyr Glu Ile Lys Val Ala Ala Glu Asn Ser Ala Gly Ile Gly			
125 130 135 140			
Val Phe Ser Asp Pro Phe Leu Phe Gln Thr Ala Glu Ser Ala Pro			
140 145 150 155			
Gly Lys Val Val Asn Leu Thr Val Glu Ala Tyr Asn Ala Ser Ala			
155 160 165 170			
Val Lys Leu Ile Trp Tyr Leu Pro Arg Gln Pro Asn Gly Lys Ile			
170 175 180 185			
Thr Ser Phe Lys Ile Ser Val Lys His Ala Arg Ser Gly Ile Val			
185 190 195 200			
Val Lys Asp Val Ser Ile Arg Val Glu Asp Ile Leu Thr Gly Lys			
200 205 210 215			
Leu Pro Glu Cys Asn Glu Asn Ser Glu Ser Phe Leu Trp Ser Thr			
215 220 225 230			
Ala Ser Pro Ser Pro Thr Leu Gly Arg Val Thr Pro Pro Ser Arg			
230 235 240 245			
Thr Thr His Ser Ser Ser Thr Leu Thr Gln Asn Glu Ile Ser Ser			
245 250 255 260			
Val Trp Lys Glu Pro Ile Ser Phe Val Val Thr His Leu Arg Pro			
260 265 270 275			
Tyr Thr Thr Tyr Leu Phe Glu Val Ser Ala Ala Thr Thr Glu Ala			
275 280 285 290			
Gly Tyr Ile Asp Ser Thr Ile Val Arg Thr Pro Glu Ser Val Pro			
290 295 300 305			
Glu Gly Pro Pro Gln Asn Cys Val Thr Gly Asn Ile Thr Gly Lys			
305 310 315 320			
Ser Phe Ser Ile Leu Trp Asp Pro Pro Thr Ile Val Thr Gly Lys			
320 325 330 335			
Phe Ser Tyr Arg Val Glu Leu Tyr Gly Pro Ser Gly Ala Gly Arg			
335 340 345 350			
Ile Leu Asp Asn Ser Thr Lys Asp Leu Lys Phe Ala Phe Thr Asn			
350 355 360 365			
Leu Thr Pro Phe Thr Met Tyr Asp Val Tyr Ile Ala Ala Glu Thr			

	365		370		375
Ser Ala Gly Thr	Gly Pro Lys Ser Asn	Ile Ser Val Phe Thr	Pro		
380	385	385	390		
Pro Asp Val Pro	Gly Ala Val Phe Asp	Leu Gln Leu Ala Glu	Val		
395	400	400	405		
Glu Ser Thr Gln	Val Arg Ile Thr Trp	Lys Lys Pro Arg Gln	Pro		
410	415	415	420		
Asn Gly Ile Ile	Asn Gln Tyr Arg Val	Lys Val Leu Val Pro	Glu		
425	430	430	435		
Thr Gly Ile Ile	Leu Glu Asn Thr Leu	Leu Thr Gly Asn Asn	Glu		
440	445	445	450		
Ile Asn Asp Pro	Met Ala Pro Glu Ile	Val Asn Ile Val Gln	Pro		
455	460	460	465		
Met Val Gly Leu	Tyr Glu Gly Ser Ala	Glu Met Ser Ser Asp	Leu		
470	475	475	480		
His Ser Leu Ala	Thr Phe Ile Tyr Asn	Ser His Pro Asp Lys	Asn		
485	490	490	495		
Phe Pro Ala Arg	Asn Arg Ala Glu Asp	Gln Thr Ser Pro Val	Val		
500	505	505	510		
Thr Thr Arg Asn	Gln Tyr Ile Thr Asp	Ile Ala Ala Glu Gln	Leu		
515	520	520	525		
Ser Tyr Val Ile	Arg Arg Leu Val Pro	Phe Thr Glu His Met	Ile		
530	535	535	540		
Ser Val Ser Ala	Phe Thr Ile Met Gly	Glu Gly Pro Pro Thr	Val		
545	550	550	555		
Leu Ser Val Arg	Thr Arg Gln Gln Val	Pro Ser Ser Ile Lys	Ile		
560	565	565	570		
Ile Asn Tyr Lys	Asn Ile Ser Ser Ser	Ser Ile Leu Leu Tyr	Trp		
575	580	580	585		
Asp Pro Pro Glu	Tyr Pro Asn Gly Lys	Ile Thr His Tyr Thr	Ile		
590	595	595	600		
Tyr Ala Met Glu	Leu Asp Thr Asn Arg	Ala Phe Gln Ile Thr	Thr		
605	610	610	615		
Ile Asp Asn Ser	Phe Leu Ile Thr Gly	Arg Lys Gln Trp Leu	Lys		
620	625	625	630		
Lys Tyr Thr Lys	Tyr Lys Met Arg Val	Ala Ala Ser Thr His	Val		
635	640	640	645		
Gly Glu Ser Ser	Leu Ser Glu Glu Asn	Asp Ile Phe Val Arg	Thr		
650	655	655	660		
Ser Glu Asp Glu	Pro Glu Ser Ser Pro	Gln Asp Val Glu Val	Ile		
665	670	670	675		
Asp Val Thr Ala	Asp Glu Ile Arg Leu	Lys Trp Ser Pro Pro	Glu		
680	685	685	690		
Lys Pro Asn Gly	Ile Ile Ile Ala Tyr	Glu Val Leu Tyr Lys	Asn		
695	700	700	705		
Ile Asp Thr Leu	Tyr Met Lys Asn Thr	Ser Thr Thr Asp Ile	Ile		
710	715	715	720		
Leu Arg Asn Leu	Arg Pro His Thr Leu	Tyr Asn Ile Ser Val	Arg		
725	730	730	735		
Ser Tyr Thr Arg	Phe Gly His Gly Asn	Gln Val Ser Ser Leu	Leu		
740	745	745	750		
Ser Val Arg Thr	Ser Glu Thr Val Pro	Asp Ser Ala Pro Glu	Asn		
755	760	760	765		
Ile Thr Tyr Lys	Asn Ile Ser Ser Gly	Glu Ile Glu Leu Ser	Phe		
770	775	775	780		
Leu Pro Pro Ser	Ser Pro Asn Gly Ile	Ile Gln Lys Tyr Thr	Ile		
785	790	790	795		
Tyr Leu Lys Arg	Ser Asn Gly Asn Glu	Glu Arg Thr Ile Asn	Thr		
800	805	805	810		
Thr Ser Leu Thr	Gln Asn Ile Lys Gly	Leu Lys Lys Tyr Thr	Gln		
815	820	820	825		
Tyr Ile Ile Glu	Val Ser Ala Ser Thr	Leu Lys Gly Glu Gly	Val		
830	835	835	840		

Arg Ser Ala Pro Ile Ser Ile Leu Thr Glu Glu Asp Ala Pro Asp
 845 850 855
 Ser Pro Pro Gln Asp Phe Ser Val Lys Gln Leu Ser Gly Val Thr
 860 865 870
 Val Lys Leu Ser Trp Gln Pro Pro Leu Glu Pro Asn Gly Ile Ile
 875 880 885
 Leu Tyr Tyr Thr Val Tyr Val Trp Arg Asn Arg Ser Ser Leu Lys
 890 895 900
 Thr Ile Asn Val Thr Glu Thr Ser Leu Glu Leu Ser Asp Leu Asp
 905 910 915
 Tyr Asn Val Glu Tyr Ser Ala Tyr Val Thr Ala Ser Thr Arg Phe
 920 925 930
 Gly Asp Gly Lys Thr Arg Ser Asn Ile Ile Ser Phe Gln Thr Pro
 935 940 945
 Glu Gly Pro Ser Asp Pro Pro Lys Asp Val Tyr Tyr Ala Asn Leu
 950 955 960
 Ser Ser Ser Ser Ile Ile Leu Phe Trp Thr Pro Pro Ser Lys Pro
 965 970 975
 Asn Gly Ile Ile Gln Tyr Tyr Ser Val Tyr Tyr Arg Asn Thr Ser
 980 985 990
 Gly Thr Phe Met Gln Asn Phe Thr Leu His Glu Val Thr Asn Asp
 995 1000 1005
 Phe Asp Asn Met Thr Val Ser Thr Ile Ile Asp Lys Leu Thr Ile
 1010 1015 1020
 Phe Ser Tyr Tyr Thr Phe Trp Leu Thr Ala Ser Thr Ser Val Gly
 1025 1030 1035
 Asn Gly Asn Lys Ser Ser Asp Ile Ile Glu Val Tyr Thr Asp Gln
 1040 1045 1050
 Asp Val Pro Glu Gly Phe Val Gly Asn Leu Thr Tyr Glu Ser Ile
 1055 1060 1065
 Ser Ser Thr Ala Ile Asn Val Ser Trp Val Pro Pro Ala Gln Pro
 1070 1075 1080
 Asn Gly Leu Val Phe Tyr Tyr Val Ser Leu Ile Leu Gln Gln Thr
 1085 1090 1095
 Pro Arg His Val Arg Pro Pro Leu Val Thr Tyr Glu Arg Ser Ile
 1100 1105 1110
 Tyr Phe Asp Asn Leu Glu Lys Tyr Thr Asp Tyr Ile Leu Lys Ile
 1115 1120 1125
 Thr Pro Ser Thr Glu Lys Gly Phe Ser Asp Thr Tyr Thr Ala Gln
 1130 1135 1140
 Leu Tyr Ile Lys Thr Glu Glu Asp Val Pro Glu Thr Ser Pro Ile
 1145 1150 1155
 Ile Asn Thr Phe Lys Asn Leu Ser Ser Thr Ser Val Leu Leu Ser
 1160 1165 1170
 Trp Asp Pro Pro Val Lys Pro Asn Gly Ala Ile Ile Ser Tyr Asp
 1175 1180 1185
 Leu Thr Leu Gln Gly Pro Asn Glu Asn Tyr Ser Phe Ile Thr Ser
 1190 1195 1200
 Asp Asn Tyr Ile Ile Leu Glu Glu Leu Ser Pro Phe Thr Leu Tyr
 1205 1210 1215
 Ser Phe Phe Ala Ala Ala Arg Thr Arg Lys Gly Leu Gly Pro Ser
 1220 1225 1230
 Ser Ile Leu Phe Phe Tyr Thr Asp Glu Ser Val Pro Leu Ala Pro
 1235 1240 1245
 Pro Gln Asn Leu Thr Leu Ile Asn Cys Thr Ser Asp Phe Val Trp
 1250 1255 1260
 Leu Lys Trp Ser Pro Ser Pro Leu Pro Gly Gly Ile Val Lys Val
 1265 1270 1275
 Tyr Ser Phe Lys Ile His Glu His Glu Thr Asp Thr Ile Tyr Tyr
 1280 1285 1290
 Lys Asn Ile Ser Gly Phe Lys Thr Glu Ala Lys Leu Val Gly Leu
 1295 1300 1305
 Glu Pro Val Ser Thr Tyr Ser Ile Arg Val Ser Ala Phe Thr Lys

1310	1315	1320
Val Gly Asn Gly Asn Gln Phe Ser Asn Val	Val Lys Phe Thr Thr	
1325	1330	1335
Gln Glu Ser Val Pro Asp Val Val Gln Asn Met Gln Cys Met Ala		
1340	1345	1350
Thr Ser Trp Gln Ser Val Leu Val Lys Trp Asp Pro Pro Lys Lys		
1355	1360	1365
Ala Asn Gly Ile Ile Thr Gln Tyr Met Val Thr Val Glu Arg Asn		
1370	1375	1380
Ser Thr Lys Val Ser Pro Gln Asp His Met Tyr Thr Phe Ile Lys		
1385	1390	1395
Leu Leu Ala Asn Thr Ser Tyr Val Phe Lys Val Arg Ala Ser Thr		
1400	1405	1410
Ser Ala Gly Glu Gly Asp Glu Ser Thr Cys His Val Ser Thr Leu		
1415	1420	1425
Pro Glu Thr Val Pro Ser Val Pro Thr Asn Ile Ala Phe Ser Asp		
1430	1435	1440
Val Gln Ser Thr Ser Ala Thr Leu Thr Trp Ile Arg Pro Asp Thr		
1445	1450	1455
Ile Leu Gly Tyr Phe Gln Asn Tyr Lys Ile Thr Thr Gln Leu Arg		
1460	1465	1470
Ala Gln Lys Cys Lys Glu Trp Glu Ser Glu Glu Cys Val Glu Tyr		
1475	1480	1485
Gln Lys Ile Gln Tyr Leu Tyr Glu Ala His Leu Thr Glu Glu Thr		
1490	1495	1500
Val Tyr Gly Leu Lys Lys Phe Arg Trp Tyr Arg Phe Gln Val Ala		
1505	1510	1515
Ser Ser Thr Asn Ala Gly Tyr Gly Asn Ala Ser Asn Trp Ile Ser		
1520	1525	1530
Thr Lys Thr Leu Pro Gly Pro Pro Asp Gly Pro Pro Glu Asn Val		
1535	1540	1545
His Val Val Ala Thr Ser Pro Phe Ser Ile Ser Ile Ser Trp Ser		
1550	1555	1560
Glu Pro Ala Val Ile Thr Gly Pro Thr Cys Tyr Leu Ile Asp Val		
1565	1570	1575
Lys Ser Val Asp Asn Asp Glu Phe Asn Ile Ser Phe Ile Lys Ser		
1580	1585	1590
Asn Glu Glu Asn Lys Thr Ile Glu Ile Lys Asp Leu Glu Ile Phe		
1595	1600	1605
Thr Arg Tyr Ser Val Val Ile Thr Ala Phe Thr Gly Asn Ile Ser		
1610	1615	1620
Ala Ala Tyr Val Glu Gly Lys Ser Ser Ala Glu Met Ile Val Thr		
1625	1630	1635
Thr Leu Glu Ser Ala Pro Lys Asp Pro Pro Asn Asn Met Thr Phe		
1640	1645	1650
Gln Lys Ile Pro Asp Glu Val Thr Lys Phe Gln Leu Thr Phe Leu		
1655	1660	1665
Pro Pro Ser Gln Pro Asn Gly Asn Ile Gln Val Tyr Gln Ala Leu		
1670	1675	1680
Val Tyr Arg Glu Asp Asp Pro Thr Ala Val Gln Ile His Asn Leu		
1685	1690	1695
Ser Ile Ile Gln Lys Thr Asn Thr Phe Val Ile Ala Met Leu Glu		
1700	1705	1710
Gly Leu Lys Gly Gly His Thr Tyr Asn Ile Ser Val Tyr Ala Val		
1715	1720	1725
Asn Ser Ala Gly Ala Gly Pro Lys Val Pro Met Arg Ile Thr Met		
1730	1735	1740
Asp Ile Lys Ala Pro Ala Arg Pro Lys Thr Lys Pro Thr Pro Ile		
1745	1750	1755
Tyr Asp Ala Thr Gly Lys Leu Leu Val Thr Ser Thr Thr Ile Thr		
1760	1765	1770
Ile Arg Met Pro Ile Cys Tyr Tyr Ser Asp Asp His Gly Pro Ile		
1775	1780	1785

Lys Asn Val Gln Val Leu Val Thr Glu Thr Gly Ala Gln His Asp
 1790 1795 1800
 Gly Asn Val Thr Lys Trp Tyr Asp Ala Tyr Phe Asn Lys Ala Arg
 1805 1810 1815
 Pro Tyr Phe Thr Asn Glu Gly Phe Pro Asn Pro Pro Cys Thr Glu
 1820 1825 1830
 Gly Lys Thr Lys Phe Ser Gly Asn Glu Glu Ile Tyr Ile Ile Gly
 1835 1840 1845
 Ala Asp Asn Ala Cys Met Ile Pro Gly Asn Glu Asp Lys Ile Cys
 1850 1855 1860
 Asn Gly Pro Leu Lys Pro Lys Lys Gln Tyr Leu Phe Lys Phe Arg
 1865 1870 1875
 Ala Thr Asn Ile Met Gly Gln Phe Thr Asp Ser Asp Tyr Ser Asp
 1880 1885 1890
 Pro Val Lys Thr Leu Gly Glu Gly Leu Ser Glu Arg Thr Val Glu
 1895 1900 1905
 Ile Ile Leu Ser Val Thr Leu Cys Ile Leu Ser Ile Ile Leu Leu
 1910 1915 1920
 Gly Thr Ala Ile Phe Ala Phe Ala Arg Ile Arg Gln Lys Gln Lys
 1925 1930 1935
 Glu Gly Gly Thr Tyr Ser Pro Gln Asp Ala Glu Ile Ile Asp Thr
 1940 1945 1950
 Lys Leu Lys Leu Asp Gln Leu Ile Thr Val Ala Asp Leu Glu Leu
 1955 1960 1965
 Lys Asp Glu Arg Leu Thr Arg Leu Leu Ser Tyr Arg Lys Ser Ile
 1970 1975 1980
 Lys Pro Ile Ser Lys Lys Ser Phe Leu Gln His Val Glu Glu Leu
 1985 1990 1995
 Cys Thr Asn Asn Asn Leu Lys Phe Gln Glu Glu Phe Ser Glu Leu
 2000 2005 2010
 Pro Lys Phe Leu Gln Asp Leu Ser Ser Thr Asp Ala Asp Leu Pro
 2015 2020 2025
 Trp Asn Arg Ala Lys Asn Arg Phe Pro Asn Ile Lys Pro Tyr Asn
 2030 2035 2040
 Asn Asn Arg Val Lys Leu Ile Ala Asp Ala Ser Val Pro Gly Ser
 2045 2050 2055
 Asp Tyr Ile Asn Ala Ser Tyr Ile Ser Gly Tyr Leu Cys Pro Asn
 2060 2065 2070
 Glu Phe Ile Ala Thr Gln Gly Pro Leu Pro Gly Thr Val Gly Asp
 2075 2080 2085
 Phe Trp Arg Met Val Trp Glu Thr Arg Ala Lys Thr Leu Val Met
 2090 2095 2100
 Leu Thr Gln Cys Phe Glu Lys Gly Arg Ile Arg Cys His Gln Tyr
 2105 2110 2115
 Trp Pro Glu Asp Asn Lys Pro Val Thr Val Phe Gly Asp Ile Val
 2120 2125 2130
 Ile Thr Lys Leu Met Glu Asp Val Gln Ile Asp Trp Thr Ile Arg
 2135 2140 2145
 Asp Leu Lys Ile Glu Arg His Gly Asp Cys Met Thr Val Arg Gln
 2150 2155 2160
 Cys Asn Phe Thr Ala Trp Pro Glu His Gly Val Pro Glu Asn Ser
 2165 2170 2175
 Ala Pro Leu Ile His Phe Val Lys Leu Val Arg Ala Ser Arg Ala
 2180 2185 2190
 His Asp Thr Thr Pro Met Ile Val His Cys Ser Ala Gly Val Gly
 2195 2200 2205
 Arg Thr Gly Val Phe Ile Ala Leu Asp His Leu Thr Gln His Ile
 2210 2215 2220
 Asn Asp His Asp Phe Val Asp Ile Tyr Gly Leu Val Ala Glu Leu
 2225 2230 2235
 Arg Ser Glu Arg Met Cys Met Val Gln Asn Leu Ala Gln Tyr Ile
 2240 2245 2250
 Phe Leu His Gln Cys Ile Leu Asp Leu Leu Ser Asn Lys Gly Ser

2255	2260	2265
Asn Gln Pro Ile Cys Phe Val Asn Tyr Ser Ala Leu Gln Lys Met		
2270	2275	2280
Asp Ser Leu Asp Ala Met Glu Gly Asp Val Glu Leu Glu Trp Glu		
2285	2290	2295
Glu Thr Thr Met		

<210> 3
 <211> 478
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7482931CD1

<400> 3		
Met Ser Gly Gly Glu Gln Leu Asp Ile Leu Ser Val Gly Ile		
1 5 10 15		
Leu Val Lys Glu Arg Trp Lys Val Leu Arg Lys Ile Gly Gly		
20 25 30		
Gly Phe Gly Glu Ile Tyr Asp Ala Leu Asp Met Leu Thr Arg Glu		
35 40 45		
Asn Val Ala Leu Lys Val Glu Ser Ala Gln Gln Pro Lys Gln Val		
50 55 60		
Leu Lys Met Glu Val Ala Val Leu Lys Lys Leu Gln Gly Lys Asp		
65 70 75		
His Val Cys Arg Phe Ile Gly Cys Gly Arg Asn Asp Arg Phe Asn		
80 85 90		
Tyr Val Val Met Gln Leu Gln Gly Arg Asn Leu Ala Asp Leu Arg		
95 100 105		
Arg Ser Gln Ser Arg Gly Thr Phe Thr Ile Ser Thr Thr Leu Arg		
110 115 120		
Leu Gly Arg Gln Ile Leu Glu Ser Ile Glu Ser Ile His Ser Val		
125 130 135		
Gly Phe Leu His Arg Asp Ile Lys Pro Ser Asn Phe Ala Met Gly		
140 145 150		
Arg Phe Pro Ser Thr Cys Arg Lys Cys Tyr Met Leu Asp Phe Gly		
155 160 165		
Leu Ala Arg Gln Phe Thr Asn Ser Cys Gly Asp Val Arg Pro Pro		
170 175 180		
Arg Ala Val Ala Gly Phe Arg Gly Thr Val Arg Tyr Ala Ser Ile		
185 190 195		
Asn Ala His Arg Asn Arg Glu Met Gly Arg His Asp Asp Leu Trp		
200 205 210		
Ser Leu Phe Tyr Met Leu Val Glu Phe Val Val Gly Gln Leu Pro		
215 220 225		
Trp Lys Lys Ile Lys Asp Lys Glu Gln Val Gly Ser Ile Lys Glu		
230 235 240		
Arg Tyr Asp His Arg Leu Met Leu Lys His Leu Pro Pro Glu Phe		
245 250 255		
Ser Ile Phe Leu Asp His Ile Ser Ser Leu Asp Tyr Phe Thr Lys		
260 265 270		
Pro Asp Tyr Gln Leu Leu Thr Ser Val Phe Asp Asn Ser Ile Lys		
275 280 285		
Thr Phe Gly Val Ile Glu Ser Asp Pro Phe Asp Trp Glu Lys Thr		
290 295 300		
Gly Asn Asp Gly Ser Leu Thr Thr Thr Thr Ser Thr Thr Pro		
305 310 315		
Gln Leu His Thr Arg Leu Thr Pro Ala Ala Ile Gly Ile Ala Asn		
320 325 330		
Ala Thr Pro Ile Pro Gly Asp Leu Leu Arg Glu Asn Thr Asp Glu		

335	340	345
Val Phe Pro Asp Glu Gln Leu Ser Asp Gly Glu Asn Gly Ile Pro	350	355
350	355	360
Val Gly Val Ser Pro Asp Lys Leu Pro Gly Ser Leu Gly His Pro	365	370
365	370	375
Arg Pro Gln Glu Lys Asp Val Trp Glu Glu Met Asp Ala Asn Lys	380	385
380	385	390
Asn Lys Ile Lys Leu Gly Ile Cys Lys Ala Ala Thr Glu Glu Glu	395	400
395	400	405
Asn Ser His Gly Gln Ala Asn Gly Leu Leu Asn Ala Pro Ser Leu	410	415
410	415	420
Gly Ser Pro Ile Arg Val Arg Ser Glu Ile Thr Gln Pro Asp Arg	425	430
425	430	435
Asp Ile Pro Leu Val Arg Lys Leu Arg Ser Ile His Ser Phe Glu	440	445
440	445	450
Leu Glu Lys Arg Leu Thr Leu Glu Pro Lys Pro Asp Thr Asp Lys	455	460
455	460	465
Phe Leu Glu Thr Trp Tyr Lys Ile Val Tyr Phe Ser Phe	470	475

<210> 4
<211> 1867
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2080788CD1

<400> 4				
Met Ala Arg Leu Ala Asp Tyr Phe Val Leu Val Ala Phe Gly Pro	1	5	10	15
His Pro Arg Gly Ser Gly Glu Gly Gln Gly Gln Ile Leu Gln Arg	20	25	30	
Phe Pro Glu Lys Asp Trp Glu Asp Asn Pro Phe Pro Gln Gly Ile	35	40	45	
Glu Leu Phe Cys Gln Pro Ser Gly Trp Gln Leu Cys Pro Glu Arg	50	55	60	
Asn Pro Pro Thr Phe Phe Val Ala Val Leu Thr Asp Ile Asn Ser	65	70	75	
Glu Arg His Tyr Cys Ala Cys Leu Thr Phe Trp Glu Pro Ala Glu	80	85	90	
Pro Ser Gln Glu Thr Thr Arg Val Glu Asp Ala Thr Glu Arg Glu	95	100	105	
Glu Glu Gly Asp Glu Gly Gly Gln Thr His Leu Ser Pro Thr Ala	110	115	120	
Pro Ala Pro Ser Ala Gln Leu Phe Ala Pro Lys Thr Leu Val Leu	125	130	135	
Val Ser Arg Leu Asp His Thr Glu Val Phe Arg Asn Ser Leu Gly	140	145	150	
Leu Ile Tyr Ala Ile His Val Glu Gly Leu Asn Val Cys Leu Glu	155	160	165	
Asn Val Ile Gly Asn Leu Leu Thr Cys Thr Val Pro Leu Ala Gly	170	175	180	
Gly Ser Gln Arg Thr Ile Ser Leu Gly Ala Gly Asp Arg Gln Val	185	190	195	
Ile Gln Thr Pro Leu Ala Asp Ser Leu Pro Val Ser Arg Cys Ser	200	205	210	
Val Ala Leu Leu Phe Arg Gln Leu Gly Ile Thr Asn Val Leu Ser	215	220	225	
Leu Phe Cys Ala Ala Leu Thr Glu His Lys Val Leu Phe Leu Ser	230	235	240	
Arg Ser Tyr Gln Arg Leu Ala Asp Ala Cys Arg Gly Leu Leu Ala				

	245	250	255											
Leu	Leu	Phe	Pro	Leu	Arg	Tyr	Ser	Phe	Thr	Tyr	Val	Pro	Ile	Leu
	260				265				265					270
Pro	Ala	Gln	Leu	Leu	Glu	Val	Leu	Ser	Thr	Pro	Thr	Pro	Phe	Ile
	275					275			280					285
Ile	Gly	Val	Asn	Ala	Ala	Phe	Gln	Ala	Glu	Thr	Gln	Glu	Leu	Leu
	290					290			295					300
Asp	Val	Ile	Val	Ala	Asp	Leu	Asp	Gly	Gly	Thr	Val	Thr	Ile	Pro
	305					305			310					315
Glu	Cys	Val	His	Ile	Pro	Pro	Leu	Pro	Glu	Pro	Leu	Gln	Ser	Gln
	320					320			325					330
Thr	His	Ser	Val	Leu	Ser	Met	Val	Leu	Asp	Pro	Glu	Leu	Glu	Leu
	335					335			340					345
Ala	Asp	Leu	Ala	Phe	Pro	Pro	Pro	Thr	Thr	Ser	Thr	Ser	Ser	Leu
	350					350			355					360
Lys	Met	Gln	Asp	Lys	Glu	Leu	Arg	Ala	Val	Phe	Leu	Arg	Leu	Phe
	365					365			370					375
Ala	Gln	Leu	Leu	Gln	Gly	Tyr	Arg	Trp	Cys	Leu	His	Val	Val	Arg
	380					380			385					390
Ile	His	Pro	Glu	Pro	Val	Ile	Arg	Phe	His	Lys	Ala	Ala	Phe	Leu
	395					395			400					405
Gly	Gln	Arg	Gly	Leu	Val	Glu	Asp	Asp	Phe	Leu	Met	Lys	Val	Leu
	410					410			415					420
Glu	Gly	Met	Ala	Phe	Ala	Gly	Phe	Val	Ser	Glu	Arg	Gly	Val	Pro
	425					425			430					435
Tyr	Arg	Pro	Thr	Asp	Leu	Phe	Asp	Glu	Leu	Val	Ala	His	Glu	Val
	440					440			445					450
Ala	Arg	Met	Arg	Ala	Asp	Glu	Asn	His	Pro	Gln	Arg	Val	Leu	Arg
	455					455			460					465
His	Val	Gln	Glu	Leu	Ala	Glu	Gln	Leu	Tyr	Lys	Asn	Glu	Asn	Pro
	470					470			475					480
Tyr	Pro	Ala	Val	Ala	Met	His	Lys	Val	Gln	Arg	Pro	Gly	Glu	Ser
	485					485			490					495
Ser	His	Leu	Arg	Arg	Val	Pro	Arg	Pro	Phe	Pro	Arg	Leu	Asp	Glu
	500					500			505					510
Gly	Thr	Val	Gln	Trp	Ile	Val	Asp	Gln	Ala	Ala	Ala	Lys	Met	Gln
	515					515			520					525
Gly	Ala	Pro	Pro	Ala	Val	Lys	Ala	Glu	Arg	Arg	Thr	Thr	Val	Pro
	530					530			535					540
Ser	Gly	Pro	Pro	Met	Thr	Ala	Ile	Leu	Glu	Arg	Cys	Ser	Gly	Leu
	545					545			550					555
His	Val	Asn	Ser	Ala	Arg	Arg	Leu	Glu	Val	Val	Arg	Asn	Cys	Ile
	560					560			565					570
Ser	Tyr	Val	Phe	Glu	Gly	Lys	Met	Leu	Glu	Ala	Lys	Lys	Leu	Leu
	575					575			580					585
Pro	Ala	Val	Leu	Arg	Ala	Leu	Lys	Gly	Arg	Ala	Ala	Arg	Arg	Cys
	590					590			595					600
Leu	Ala	Gln	Glu	Leu	His	Leu	His	Val	Gln	Gln	Asn	Arg	Ala	Val
	605					605			610					615
Leu	Asp	His	Gln	Gln	Phe	Asp	Phe	Val	Val	Arg	Met	Met	Asn	Cys
	620					620			625					630
Cys	Leu	Gln	Asp	Cys	Thr	Ser	Leu	Asp	Glu	His	Gly	Ile	Ala	Ala
	635					635			640					645
Ala	Leu	Leu	Pro	Leu	Val	Thr	Ala	Phe	Cys	Arg	Lys	Leu	Ser	Pro
	650					650			655					660
Gly	Val	Thr	Gln	Phe	Ala	Tyr	Ser	Cys	Val	Gln	Glu	His	Val	Val
	665					665			670					675
Trp	Ser	Thr	Pro	Gln	Phe	Trp	Glu	Ala	Met	Phe	Tyr	Gly	Asp	Val
	680					680			685					690
Gln	Thr	His	Ile	Arg	Ala	Leu	Tyr	Leu	Glu	Pro	Thr	Glu	Asp	Leu
	695					695			700					705
Ala	Pro	Ala	Gln	Glu	Val	Gly	Glu	Ala	Pro	Ser	Gln	Glu	Asp	Glu
	710					710			715					720

Arg Ser Ala Leu Asp Val Ala Ser Glu Gln Arg Arg Leu Trp Pro
 725 730 735
 Thr Leu Ser Arg Glu Lys Gln Gln Glu Leu Val Gln Lys Glu Glu
 740 745 750
 Ser Thr Val Phe Ser Gln Ala Ile His Tyr Ala Asn Arg Met Ser
 755 760 765
 Tyr Leu Leu Leu Pro Leu Asp Ser Ser Lys Ser Arg Leu Leu Arg
 770 775 780
 Glu Arg Ala Gly Leu Gly Asp Leu Glu Ser Ala Ser Asn Ser Leu
 785 790 795
 Val Thr Asn Ser Met Ala Gly Ser Val Ala Glu Ser Tyr Asp Thr
 800 805 810
 Glu Ser Gly Phe Glu Asp Ala Glu Thr Cys Asp Val Ala Gly Ala
 815 820 825
 Val Val Arg Phe Ile Asn Arg Phe Val Asp Lys Val Cys Thr Glu
 830 835 840
 Ser Gly Val Thr Ser Asp His Leu Lys Gly Leu His Val Met Val
 845 850 855
 Pro Asp Ile Val Gln Met His Ile Glu Thr Leu Glu Ala Val Gln
 860 865 870
 Arg Glu Ser Arg Arg Leu Pro Pro Ile Gln Lys Pro Lys Leu Leu
 875 880 885
 Arg Pro Arg Leu Leu Pro Gly Glu Glu Cys Val Leu Asp Gly Leu
 890 895 900
 Arg Val Tyr Leu Leu Pro Asp Gly Arg Glu Glu Gly Ala Gly Gly
 905 910 915
 Ser Ala Gly Gly Pro Ala Leu Leu Pro Ala Glu Gly Ala Val Phe
 920 925 930
 Leu Thr Thr Tyr Arg Val Ile Phe Thr Gly Met Pro Thr Asp Pro
 935 940 945
 Leu Val Gly Glu Gln Val Val Val Arg Ser Phe Pro Val Ala Ala
 950 955 960
 Leu Thr Lys Glu Lys Arg Ile Ser Val Gln Thr Pro Val Asp Gln
 965 970 975
 Leu Leu Gln Asp Gly Leu Gln Leu Arg Ser Cys Thr Phe Gln Leu
 980 985 990
 Leu Lys Met Ala Phe Asp Glu Glu Val Gly Ser Asp Ser Ala Glu
 995 1000 1005
 Leu Phe Arg Lys Gln Leu His Lys Leu Arg Tyr Pro Pro Asp Ile
 1010 1015 1020
 Arg Ala Thr Phe Ala Phe Thr Leu Gly Ser Ala His Thr Pro Gly
 1025 1030 1035
 Arg Pro Pro Arg Val Thr Lys Asp Lys Gly Pro Ser Leu Arg Thr
 1040 1045 1050
 Leu Ser Arg Asn Leu Val Lys Asn Ala Lys Lys Thr Ile Gly Arg
 1055 1060 1065
 Gln His Val Thr Arg Lys Lys Tyr Asn Pro Pro Ser Trp Glu His
 1070 1075 1080
 Arg Gly Gln Pro Pro Glu Asp Gln Glu Asp Glu Ile Ser Val
 1085 1090 1095
 Ser Glu Glu Leu Glu Pro Ser Thr Leu Thr Pro Ser Ser Ala Leu
 1100 1105 1110
 Lys Pro Ser Asp Arg Met Thr Met Ser Ser Leu Val Glu Arg Ala
 1115 1120 1125
 Cys Cys Arg Asp Tyr Gln Arg Leu Gly Leu Gly Thr Leu Ser Ser
 1130 1135 1140
 Ser Leu Ser Arg Ala Lys Ser Glu Pro Phe Arg Ile Ser Pro Val
 1145 1150 1155
 Asn Arg Met Tyr Ala Ile Cys Arg Ser Tyr Pro Gly Leu Leu Ile
 1160 1165 1170
 Val Pro Gln Ser Val Gln Asp Asn Ala Leu Gln Arg Val Ser Arg
 1175 1180 1185
 Cys Tyr Arg Gln Asn Arg Phe Pro Val Val Cys Trp Arg Ser Gly

1190	1195	1200
Arg Ser Lys Ala Val Leu Leu Arg Ser Gly Gly Leu His Gly Lys	1205	1210
Gly Val Val Gly Leu Phe Lys Ala Gln Asn Ala Pro Ser Pro Gly	1220	1225
Gln Ser Gln Ala Asp Ser Ser Ser Leu Glu Gln Glu Lys Tyr Leu	1235	1240
Gln Ala Val Val Ser Ser Met Pro Arg Tyr Ala Asp Ala Ser Gly	1250	1255
Arg Asn Thr Leu Ser Gly Phe Ser Ser Ala His Met Gly Ser His	1265	1270
Gly Lys Trp Gly Ser Val Arg Thr Ser Gly Arg Ser Ser Gly Leu	1280	1285
Gly Thr Asp Val Gly Ser Arg Leu Ala Gly Arg Asp Ala Leu Ala	1295	1300
Pro Pro Gln Ala Asn Gly Gly Pro Pro Asp Pro Gly Phe Leu Arg	1310	1315
Pro Gln Arg Ala Ala Leu Tyr Ile Leu Gly Asp Lys Ala Gln Leu	1325	1330
Lys Gly Val Arg Ser Asp Pro Leu Gln Gln Trp Glu Leu Val Pro	1340	1345
Ile Glu Val Phe Glu Ala Arg Gln Val Lys Ala Ser Phe Lys Lys	1355	1360
Leu Leu Lys Ala Cys Val Pro Gly Cys Pro Ala Ala Glu Pro Ser	1370	1375
Pro Ala Ser Phe Leu Arg Ser Leu Glu Asp Ser Glu Trp Leu Ile	1385	1390
Gln Ile His Lys Leu Leu Gln Val Ser Val Leu Val Val Glu Leu	1400	1405
Leu Asp Ser Gly Ser Ser Val Leu Val Gly Leu Glu Asp Gly Trp	1415	1420
Asp Ile Thr Thr Gln Val Val Ser Leu Val Gln Leu Leu Ser Asp	1430	1435
Pro Phe Tyr Arg Thr Leu Glu Gly Phe Arg Leu Leu Val Glu Lys	1445	1450
Glu Trp Leu Ser Phe Gly His Arg Phe Ser His Arg Gly Ala His	1460	1465
Thr Leu Ala Gly Gln Ser Ser Gly Phe Thr Pro Val Phe Leu Gln	1475	1480
Phe Leu Asp Cys Val His Gln Val His Leu Gln Phe Pro Met Glu	1490	1495
Phe Glu Phe Ser Gln Phe Tyr Leu Lys Phe Leu Gly Tyr His His	1505	1510
Val Ser Arg Arg Phe Arg Thr Phe Leu Leu Asp Ser Asp Tyr Glu	1520	1525
Arg Ile Glu Leu Gly Leu Leu Tyr Glu Glu Lys Gly Glu Arg Arg	1535	1540
Gly Gln Val Pro Cys Arg Ser Val Trp Glu Tyr Val Asp Arg Leu	1550	1555
Ser Lys Arg Thr Pro Val Phe His Asn Tyr Met Tyr Ala Pro Glu	1565	1570
Asp Ala Glu Val Leu Arg Pro Tyr Ser Asn Val Ser Asn Leu Lys	1580	1585
Val Trp Asp Phe Tyr Thr Glu Glu Thr Leu Ala Glu Gly Pro Pro	1595	1600
Tyr Asp Trp Glu Leu Ala Gln Gly Pro Pro Glu Pro Pro Glu Glu	1610	1615
Glu Arg Ser Asp Gly Gly Ala Pro Gln Ser Arg Arg Arg Val Val	1625	1630
Trp Pro Cys Tyr Asp Ser Cys Pro Arg Ala Gln Pro Asp Ala Ile	1640	1645
Ser Arg Leu Leu Glu Leu Gln Arg Leu Glu Thr Glu Leu Gly	1655	1660
		1665

Gln Pro Ala Glu Arg Trp Lys Asp Thr Trp Asp Arg Val Lys Ala
 1670 1675 1680
 Ala Gln Arg Leu Glu Gly Arg Pro Asp Gly Arg Gly Thr Pro Ser
 1685 1690 1695
 Ser Leu Leu Val Ser Thr Ala Pro His His Arg Arg Ser Leu Gly
 1700 1705 1710
 Val Tyr Leu Gln Glu Gly Pro Val Gly Ser Thr Leu Ser Leu Ser
 1715 1720 1725
 Leu Asp Ser Asp Gln Ser Ser Gly Ser Thr Thr Ser Gly Ser Arg
 1730 1735 1740
 Gln Ala Ala Arg Arg Ser Thr Ser Thr Leu Tyr Ser Gln Phe Gln
 1745 1750 1755
 Thr Ala Glu Ser Glu Asn Arg Ser Tyr Glu Gly Thr Leu Tyr Lys
 1760 1765 1770
 Lys Gly Ala Phe Met Lys Pro Trp Lys Ala Arg Trp Phe Val Leu
 1775 1780 1785
 Asp Lys Thr Lys His Gln Leu Arg Tyr Tyr Asp His Arg Val Asp
 1790 1795 1800
 Thr Glu Cys Lys Gly Val Ile Asp Leu Ala Glu Val Glu Ala Val
 1805 1810 1815
 Ala Pro Gly Thr Pro Thr Met Gly Ala Pro Lys Thr Val Asp Glu
 1820 1825 1830
 Lys Ala Phe Phe Asp Val Lys Thr Thr Arg Arg Val Tyr Asn Phe
 1835 1840 1845
 Cys Ala Gln Asp Val Pro Ser Ala Gln Gln Trp Val Asp Arg Ile
 1850 1855 1860
 Gln Ser Cys Leu Ser Asp Ala
 1865

<210> 5
 <211> 514
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 71918969CD1

<400> 5
 Met Ala Asp Ser Gly Leu Asp Lys Lys Ser Thr Lys Cys Pro Asp
 1 5 10 15
 Cys Ser Ser Ala Ser Gln Lys Asp Val Leu Cys Val Cys Ser Ser
 20 25 30
 Lys Thr Arg Val Pro Pro Val Leu Val Val Glu Met Ser Gln Thr
 35 40 45
 Ser Ser Ile Gly Ser Ala Glu Ser Leu Ile Ser Leu Glu Arg Lys
 50 55 60
 Lys Glu Lys Asn Ile Asn Arg Asp Ile Thr Ser Arg Lys Asp Leu
 65 70 75
 Pro Ser Arg Thr Ser Asn Val Glu Arg Lys Ala Ser Gln Gln
 80 85 90
 Trp Gly Arg Gly Asn Phe Thr Glu Gly Lys Val Pro His Ile Arg
 95 100 105
 Ile Glu Asn Gly Ala Ala Ile Glu Glu Ile Tyr Thr Phe Gly Arg
 110 115 120
 Ile Leu Gly Lys Gly Ser Phe Gly Ile Val Ile Glu Ala Thr Asp
 125 130 135
 Lys Glu Thr Glu Thr Lys Trp Ala Ile Lys Lys Val Asn Lys Glu
 140 145 150
 Lys Ala Gly Ser Ser Ala Val Lys Leu Leu Glu Arg Glu Val Asn
 155 160 165
 Ile Leu Lys Ser Val Lys His Glu His Ile Ile His Leu Glu Gln
 170 175 180

Val	Phe	Glu	Thr	Pro	Lys	Lys	Met	Tyr	Leu	Val	Met	Glu	Leu	Cys
185									190					195
Glu	Asp	Gly	Glu	Leu	Lys	Glu	Ile	Leu	Asp	Arg	Lys	Gly	His	Phe
200									205					210
Ser	Glu	Asn	Glu	Thr	Arg	Trp	Ile	Ile	Gln	Ser	Leu	Ala	Ser	Ala
215									220					225
Ile	Ala	Tyr	Leu	His	Asn	Asn	Asp	Ile	Val	His	Arg	Asp	Leu	Lys
230									235					240
Leu	Glu	Asn	Ile	Met	Val	Lys	Ser	Ser	Leu	Ile	Asp	Asp	Asn	Asn
245									250					255
Glu	Ile	Asn	Leu	Asn	Ile	Lys	Val	Thr	Asp	Phe	Gly	Leu	Ala	Val
260									265					270
Lys	Lys	Gln	Ser	Arg	Ser	Glu	Ala	Met	Leu	Gln	Ala	Thr	Cys	Gly
275									280					285
Thr	Pro	Ile	Tyr	Met	Ala	Pro	Glu	Val	Ile	Ser	Ala	His	Asp	Tyr
290									295					300
Ser	Gln	Gln	Cys	Asp	Ile	Trp	Ser	Ile	Gly	Val	Val	Met	Tyr	Met
305									310					315
Leu	Leu	Arg	Gly	Glu	Pro	Pro	Phe	Leu	Ala	Ser	Ser	Glu	Glu	Lys
320									325					330
Leu	Phe	Glu	Leu	Ile	Arg	Lys	Gly	Glu	Leu	His	Phe	Glu	Asn	Ala
335									340					345
Val	Trp	Asn	Ser	Ile	Ser	Asp	Cys	Ala	Lys	Ser	Val	Leu	Lys	Gln
350									355					360
Leu	Met	Lys	Val	Asp	Pro	Ala	His	Arg	Ile	Thr	Ala	Lys	Glu	Leu
365									370					375
Leu	Asp	Asn	Gln	Trp	Leu	Thr	Gly	Asn	Lys	Leu	Ser	Ser	Val	Arg
380									385					390
Pro	Thr	Asn	Val	Leu	Glu	Met	Met	Lys	Glu	Trp	Lys	Asn	Asn	Pro
395									400					405
Glu	Ser	Val	Glu	Glu	Asn	Thr	Thr	Glu	Glu	Lys	Asn	Lys	Pro	Ser
410									415					420
Thr	Glu	Glu	Lys	Leu	Lys	Ser	Tyr	Gln	Pro	Trp	Gly	Asn	Val	Pro
425									430					435
Asp	Ala	Asn	Tyr	Thr	Ser	Asp	Glu	Glu	Glu	Glu	Lys	Gln	Ser	Thr
440									445					450
Ala	Tyr	Glu	Lys	Gln	Phe	Pro	Ala	Thr	Ser	Lys	Asp	Asn	Phe	Asp
455									460					465
Met	Cys	Ser	Ser	Ser	Phe	Thr	Ser	Ser	Lys	Leu	Leu	Pro	Ala	Glu
470									475					480
Ile	Lys	Gly	Glu	Met	Glu	Lys	Thr	Pro	Val	Thr	Pro	Ser	Gln	Gly
485									490					495
Thr	Ala	Thr	Lys	Tyr	Pro	Ala	Lys	Ser	Gly	Ala	Leu	Ser	Arg	Thr
500									505					510
Lys Lys Lys Leu														

<210> 6
<211> 338
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8187571CD1

Met	Ala	Ser	Arg	Leu	Leu	His	Arg	His	Ile	Arg	Glu	Gln	Leu	Lys
1				5					10					15
Asp	Leu	Val	Glu	Ile	Leu	Gln	Asp	Pro	Ser	Pro	Pro	Pro	Leu	Cys
									20					30
Leu	Pro	Thr	Thr	Pro	Gly	Thr	Pro	Asp	Ser	Ser	Asp	Pro	Ser	His
									35					45

Leu	Leu	Gly	Pro	Gln	Ser	Cys	Trp	Ser	Ser	Gln	Lys	Glu	Val	Ser
				50						55				60
His	Glu	Ser	Leu	Val	Val	Gly	Ala	Ile	Glu	Asn	Ala	Phe	Gln	Leu
				65					70					75
Met	Asp	Glu	Gln	Met	Ala	Arg	Glu	Arg	Arg	Gly	His	Gln	Val	Glu
				80					85					90
Gly	Gly	Cys	Cys	Ala	Leu	Val	Val	Ile	Tyr	Leu	Leu	Gly	Lys	Val
				95					100					105
Tyr	Val	Ala	Asn	Ala	Gly	Asp	Ser	Arg	Ala	Ile	Ile	Val	Arg	Asn
				110					115					120
Gly	Glu	Ile	Ile	Pro	Met	Ser	Arg	Glu	Phe	Thr	Pro	Glu	Thr	Glu
				125					130					135
Arg	Gln	Arg	Leu	Gln	Leu	Leu	Gly	Phe	Leu	Lys	Pro	Glu	Leu	Leu
				140					145					150
Gly	Ser	Glu	Phe	Thr	His	Leu	Glu	Phe	Pro	Arg	Arg	Val	Leu	Pro
				155					160					165
Lys	Glu	Leu	Gly	Gln	Arg	Met	Leu	Tyr	Arg	Asp	Gln	Asn	Met	Thr
				170					175					180
Gly	Trp	Ala	Tyr	Lys	Lys	Ile	Glu	Leu	Glu	Asp	Leu	Arg	Phe	Pro
				185					190					195
Leu	Val	Cys	Gly	Glu	Gly	Lys	Lys	Ala	Arg	Val	Met	Ala	Thr	Ile
				200					205					210
Gly	Val	Thr	Arg	Gly	Leu	Gly	Asp	His	Ser	Leu	Lys	Val	Cys	Ser
				215					220					225
Ser	Thr	Leu	Pro	Ile	Lys	Pro	Phe	Leu	Ser	Cys	Phe	Pro	Glu	Val
				230					235					240
Arg	Val	Tyr	Asp	Leu	Thr	Gln	Tyr	Glu	His	Cys	Pro	Asp	Asp	Val
				245					250					255
Leu	Val	Leu	Gly	Thr	Asp	Gly	Leu	Trp	Asp	Val	Thr	Thr	Asp	Cys
				260					265					270
Glu	Val	Ala	Ala	Thr	Val	Asp	Arg	Val	Leu	Ser	Ala	Tyr	Glu	Pro
				275					280					285
Asn	Asp	His	Ser	Arg	Tyr	Thr	Ala	Leu	Ala	Gln	Ala	Leu	Val	Leu
				290					295					300
Gly	Ala	Arg	Gly	Thr	Pro	Arg	Asp	Arg	Gly	Trp	Arg	Leu	Pro	Asn
				305					310					315
Asn	Lys	Leu	Gly	Ser	Gly	Asp	Asp	Ile	Ser	Val	Phe	Val	Ile	Pro
				320					325					330
Leu	Gly	Gly	Pro	Gly	Ser	Tyr	Ser							
				335										

<210> 7

<211> 321

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7494145CD1

<400> 7

Met	Thr	Ser	Phe	His	Pro	Arg	Gly	Leu	Gln	Ala	Ala	Arg	Ala	Gln
1				5					10					15
Lys	Phe	Lys	Ser	Lys	Arg	Pro	Arg	Ser	Asn	Ser	Asp	Cys	Phe	Gln
				20					25					30
Glu	Glu	Asp	Leu	Arg	Gln	Gly	Phe	Gln	Trp	Arg	Lys	Ser	Leu	Pro
				35					40					45
Phe	Gly	Ala	Ala	Ser	Ser	Tyr	Leu	Asn	Leu	Glu	Lys	Leu	Gly	Glu
				50					55					60
Gly	Ser	Tyr	Ala	Thr	Val	Tyr	Lys	Gly	Ile	Ser	Arg	Ile	Asn	Gly
				65					70					75
Gln	Leu	Val	Ala	Leu	Lys	Val	Ile	Ser	Met	Asn	Ala	Glu	Glu	Gly
				80					85					90

Val	Pro	Phe	Thr	Ala	Ile	Arg	Glu	Ala	Ser	Leu	Leu	Lys	Gly	Leu
					95					100				105
Lys	His	Ala	Asn	Ile	Val	Leu	Leu	His	Asp	Ile	Ile	His	Thr	Lys
					110					115				120
Glu	Thr	Leu	Thr	Phe	Val	Phe	Glu	Tyr	Met	His	Thr	Asp	Leu	Ala
					125					130				135
Gln	Tyr	Met	Ser	Gln	His	Pro	Gly	Gly	Leu	His	Pro	His	Asn	Val
					140					145				150
Arg	Leu	Phe	Met	Phe	Gln	Leu	Leu	Arg	Gly	Leu	Ala	Tyr	Ile	His
					155					160				165
His	Gln	His	Val	Leu	His	Arg	Asp	Leu	Lys	Pro	Gln	Asn	Leu	Leu
					170					175				180
Ile	Ser	His	Leu	Gly	Glu	Leu	Lys	Leu	Ala	Asp	Phe	Gly	Leu	Ala
					185					190				195
Arg	Ala	Lys	Ser	Ile	Pro	Ser	Gln	Thr	Tyr	Ser	Ser	Glu	Val	Val
					200					205				210
Thr	Leu	Trp	Tyr	Arg	Pro	Pro	Asp	Ala	Leu	Leu	Gly	Ala	Thr	Glu
					215					220				225
Tyr	Ser	Ser	Glu	Leu	Asp	Ile	Trp	Gly	Ala	Gly	Cys	Ile	Phe	Ile
					230					235				240
Glu	Met	Phe	Gln	Gly	Gln	Pro	Leu	Phe	Pro	Gly	Val	Ser	Asn	Ile
					245					250				255
Leu	Glu	Gln	Leu	Glu	Lys	Ile	Trp	Glu	Val	Leu	Gly	Val	Pro	Thr
					260					265				270
Glu	Asp	Thr	Trp	Pro	Gly	Val	Ser	Lys	Leu	Pro	Asn	Tyr	Asn	Pro
					275					280				285
Glu	Glu	Ser	Leu	Phe	Thr	Val	Ser	Gly	Val	Arg	Leu	Lys	Pro	Glu
					290					295				300
Met	Cys	Asp	Leu	Leu	Ala	Ser	Tyr	Gln	Lys	Gly	His	His	Pro	Ala
					305					310				315
Gln	Phe	Ser	Lys	Cys	Trp									
					320									

<210> 8

<211> 292

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5807954CD1

<400> 8

Met	Lys	Ala	Thr	Val	Leu	Met	Arg	Gln	Pro	Gly	Arg	Val	Gln	Glu
					1	5				10				15
Ile	Val	Gly	Ala	Leu	Arg	Lys	Gly	Gly	Gly	Asp	Arg	Leu	Gln	Val
					20					25				30
Ile	Ser	Asp	Phe	Asp	Met	Thr	Leu	Ser	Arg	Phe	Ala	Tyr	Asn	Gly
					35					40				45
Lys	Arg	Cys	Pro	Ser	Ser	Tyr	Asn	Ile	Leu	Asp	Asn	Ser	Lys	Ile
					50					55				60
Ile	Ser	Glu	Glu	Cys	Arg	Lys	Glu	Leu	Thr	Ala	Leu	Leu	His	His
					65					70				75
Tyr	Tyr	Pro	Ile	Glu	Ile	Asp	Pro	His	Arg	Thr	Val	Lys	Glu	Lys
					80					85				90
Leu	Pro	His	Met	Val	Glu	Trp	Trp	Thr	Lys	Ala	His	Asn	Leu	Leu
					95					100				105
Cys	Gln	Gln	Lys	Ile	Gln	Lys	Phe	Gln	Ile	Ala	Gln	Val	Val	Arg
					110					115				120
Glu	Ser	Asn	Ala	Met	Leu	Arg	Glu	Gly	Tyr	Lys	Thr	Phe	Phe	Asn
					125					130				135
Thr	Leu	Tyr	His	Asn	Asn	Ile	Pro	Leu	Phe	Ile	Phe	Ser	Ala	Gly
					140					145				150

Ile	Gly	Asp	Ile	Leu	Glu	Glu	Ile	Ile	Arg	Gln	Met	Lys	Val	Phe
155									160					165
His	Pro	Asn	Ile	His	Ile	Val	Ser	Asn	Tyr	Met	Asp	Phe	Asn	Glu
170									175					180
Asp	Gly	Phe	Leu	Gln	Gly	Phe	Lys	Gly	Gln	Leu	Ile	His	Thr	Tyr
185									190					195
Asn	Lys	Asn	Ser	Ser	Ala	Cys	Glu	Asn	Ser	Gly	Tyr	Phe	Gln	Gln
200									205					210
Leu	Glu	Gly	Lys	Thr	Asn	Val	Ile	Leu	Leu	Gly	Asp	Ser	Ile	Gly
215									220					225
Asp	Leu	Thr	Met	Ala	Asp	Gly	Val	Pro	Gly	Val	Gln	Asn	Ile	Leu
230									235					240
Lys	Ile	Gly	Phe	Leu	Asn	Asp	Lys	Val	Glu	Glu	Arg	Arg	Glu	Arg
245									250					255
Tyr	Met	Asp	Ser	Tyr	Asp	Ile	Val	Leu	Glu	Lys	Asp	Glu	Thr	Leu
260									265					270
Asp	Val	Val	Asn	Gly	Leu	Leu	Gln	His	Ile	Leu	Cys	Gln	Gly	Val
275									280					285
Gln	Leu	Glu	Met	Gln	Gly	Pro								
290														

<210> 9
<211> 1605
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 55074884CB1

<400> 9														
ccaagccca	gcctggtcca	cctcgaggc	ctctaggacc	cgggggcgcc	cgcgccgcgc	60								
ccggctccca	caaatacgact	cctgggggg	cgcctgagcc	cccaaaaatag	atccctcaggg	120								
cccaaaaagac	gactcttcgg	cgggcgcct	gggaccggca	gaagctgggc	gcccggggc	180								
cgccctcgccc	gtacatccac	cgttgggtcg	cgtcgccccc	tcacttcc	tcgggagcgc	240								
gcgagccgcg	ggcgcggagg	agcagctggc	gcccgcggga	gtcacgtgt	gcgtcaacgt	300								
ctccccgcag	cagcccgcc	cgcgcgcgc	cggcgtggca	gagctgcgc	tgcccggtgt	360								
cgacgaccgg	gctgaggacc	tgctggcgca	cctggagccc	acgtgcgcgc	ccatggaggc	420								
cgccgggtgcgc	gccggcgccg	cctgcctagt	ctactgcaag	aacggccgc	gccgctcgcc	480								
cgccgtctgc	accgcgtacc	tcatgcggca	ccgcggcctc	agcctggcga	aggcatttcca	540								
gatggtaaag	agcgctcgcc	cggtagcaga	accgaacccg	ggcttctgg	ctcagctcca	600								
gaagttatgag	gaggccctcc	aggcccagtc	ctgcctgcag	ggagagcccc	cagcctttagg	660								
gttggccct	gaggcttggaa	gcttgaaggc	ctgctgcctg	gaggaaggat	gtccctgcac	720								
tgatacagaa	ggctggtctt	tacccttctt	cctcaactgtc	atatcgagtt	ttccctttgtg	780								
tgtgtgtgt	tgaaacatag	tgcttttaat	tttatatttc	cggtctgggt	gtatgcacaaa	840								
ttatcttaca	gacacaaaaaa	gaaatacatt	ttgtaaaaca	tcgaagacaa	attaagaaaa	900								
agcaaatagg	gatctccat	tatttacact	ccagaatctt	gagtgtttgg	gggtgcgggc	960								
gcattctacc	agtctcaggg	aaggacatga	gtacagacta	cagtaatcag	aatctgagtc	1020								
ataggtaaaag	tcagaaaaag	caactcttct	tgtgtttatt	tttagtagt	tcataagaac	1080								
ggactagttc	ttcctgcgt	accacggat	cttctgtttt	agaaaaatgc	tacccacgt	1140								
ggacatttag	atcaagaaaag	ctgttgcgt	agacattttt	gaaggcaac	ttgggtgtatt	1200								
ggggaaattat	tctttcatc	agcccctct	caatacagct	ggagctgtcc	cataggagtt	1260								
cggacaat	cgacatctt	ttttttttt	ctttttttt	ttttttttt	gtttgtttga	1320								
gacagagtct	ctgtgtccca	ggctggagtg	cagtggcgca	atcttgctc	actgcaaccc	1380								
ccgcctccca	ggttcaagt	attcttcgtc	ctcagctcc	ttagtagct	ggactacagg	1440								
tgtgtgacac	cagacccggc	taattttttt	agtatttttta	gtagagatgg	gggggtttca	1500								
accatgttac	ccaggatgt	tcgtactctt	gacctcagag	atccagccac	ctccatcttc	1560								
aacaatcg	atgtgaggaa	aacagagtga	cgaacattcg	ggtac		1605								

<210> 10
<211> 7225
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480588CB1

aaccttcttcc ctacctcgtg tctcttatca tgggatcccc cagtaaagcc aaatggtgca 3540
ataataaagt atgatttaac ttacaaggga ccaaataaaa attattctt cattacttct 3600
gataattaca taatattgga agagcttca ccatttacat tatatagctt ttttgctgcc 3660
gcaagaacta gaaaaggact tggtccttcc agtattctt tctttacac agatgagtca 3720
gtgccgttag cacctccaca aaatttgact ttaatcaact gtacttcaga ctttgtatgg 3780
ctgaaatgga gcccaagtcc tcttcagggt gttattgtt aagtataat ttttaaaatt 3840
catgaacatg aaactgacac tatataattt aagaataat caggattta aactgaagcc 3900
aaacttggtg gactggAAC agtcagcacc tactctatcc gtgtatctgc gttcacccaa 3960
gttggaaatg gcaatcaatt tagtaatgtt gtaaaaattca caacccaaaga atcagttcca 4020
gatgtcgtgc agaataatgca gtgcattggca actagctgc agtcagttt agtgaatagg 4080
gatccaccca aaaaggcaaa tggataata acgcagtata tggtaacat taaaaaggat 4140
tctacaaaaag tttctccca agatcacatg tacactttca taaaagcttct tgccaatacc 4200
tcatatgtct taaaagtaag agcttcaacc tcaacttcgt gctcaaaaat gcaaaagaatg 4260
catgtcagca cactacctga aacagttccc agtgttccca caaatattgc ttttctgtat 4320
gttcagtcac cttagtgcac attgacatgg ataagacctg acactatcct tggctacttt 4380
caaaattaca aaattaccac tcaacttcgt gctcaaaaat gcaaaagaatg ggaatccgaa 4440
gaatgtgttg aatatcaaaa aattcaatac ctctatgaag ctcacttaac tgaagagaca 4500
gtatatggat taaaagaaatt tagatggat agattccaaag tgggttccag caccatgtct 4560
ggctatggca atgcttcaaa ctggatttct acaaaaactc tgcctggccc tccagatgg 4620
cctctgtaaa atgttcatgt agtagacaaca tcacccttca gcacatcggcat aagctggagt 4680
gaacctgctg tcattactgg accaacatgt tatctgattt atgtcaatc ggtagataat 4740
gatgaattta atatatcctt catcaagtca aatgaagaaa ataaaaccat agaaatttaaa 4800
gatttagaaa tattcacaaag gtattctgtt gtgtacactg cattacttgg gaacattatgt 4860
gctgcatacg tagaaggggaa gtcagtgct gaaatgattt ttactactt agaatacgcc 4920
ccaaaggacc cacctaacaat caattaaatg ttccttcctt cagaagatac cagatgaagt 4980
caattaaatg ttccttcctt ttctcaaccc atggaaaat tccaatgtt tcaagctctg 5040
gtttaccgag aagatgatcc tactgtgtc cagattcaca acctcgttat tatacagaaaa 5100
accaacacat tcgtcattgc aatgctagaa ggactaaaag gtggacatac atacaatatc 5160
agtgttacg cagtcaatag tgctgtgca gtcacaaagg ttccgatgat aataaccatg 5220
gatatacaaaat cttccagcacc accaaaaacc aaacccaaccc ctatttatga tgccacagga 5280
aaactgctt tgacttcaac aacaattaca atcagaatgc caatatgtt ctacagtgat 5340
gatcatggac caataaaaaaa tgcataatgtt ttgtgacatc aaacaggagc tcgtacat 5400
ggaaatgtaa caaagtggta tgcataatgtt ttaataaaag caaggccata ttttacaaat 5460
gaaggcttcc ttaacccttcc atgtacagaa gaaaagacaa agtttagtgg caatgaagaa 5520
atctacatca taggtgctga taatgcatttgc atgattccctg gcaatgaaga caaaatttgc 5580
aatggaccac tggaaaccaaa aaagaataac ttatttaat ttagagctac aaatattatg 5640
ggacaattta ctgactctga ttattctgac cttgtttaaga cttagggga aggactttca 5700
gaaagaaccg tagagatcat tctttccgtc actttgtgtt tccttcaat tggcacatc 5760
ggacacgta ttttgcatt tgcgagaatt cgacacaaagc agaaagaagg catcgttcc 5820
tctcttcagg atgcagaaaattt tattgacact aaatttgaac tggatcagct catcacatg 5880
gcagacatgg aactgaagga cgagagatta acgcgttac tttagttat aatctcttca 5940
aagccaataa gcaagaaatc cttcctgcaat catgttgaag agctttgcac aaaacaacaac 6000
ctaaaggttt aagaagaattt tcggattt caaaaatttc ttcaaggatct ttcttcaact 6060
gatgtctgatc tgcccttggaa tagagaaaaa aaccgcttcc caaacataaa accatataat 6120
aataacagag taaagctgat agctgacgct agtgttccag gttcggatta tattatgtcc 6180
agctatattt ctggtttattt atgtccaaat gaatttattt ctactcaagg tccactacca 6240
ggaacagttt gagatttttt gagaatgggt tggggaaacca gagaaaaaac attagtaatg 6300
ctaacacagt gtttggaaaaa aggacggatc agatgccatc agtattggcc agaggacaac 6360
aagccagttt ctgtctttgg agatataatg attacaaaagc taatggagga tggtaaaata 6420
gattggacta tcaggatctt gaaaattgaa aggcatgggg attgcattac tggtaaaatg 6480
tgtaacttta ctgccttggcc agagcatggg ttccctgaga acagccccc tctaatttac 6540
tttggtaagt tggttcgagc aagcaggggca catgacacca cacctatgat tggtaactt 6600
agtgtctggag ttggaaagaaac tggagttttt attgctctgg accatataac acaacatata 6660
aatgaccatg attttggaa tataatatggc ttagtagctg aactgagaag tggatcttca 6720
tgcattgtgc agaatttggc acagtataatc ttttacacc agtgcattct tggtaactt 6780
tcaaaaaagg gaatgtatca gcccattgtt tggtaactt attcagcatc tcagaagatg 6840
gactcttgg acgccccatggaa aggtgtatgtt gagcttgaat gggaaagaaac cactatgtaa 6900
atattcagac caaaggatatac aattgaaaga gattttaaa tcccaaggggc caaagttacc 6960
ccctcattct tccgaatttga aatgtgcac cttaaagaaa tatctatgtct tctctcaact 7020
tgccttcca acggatttggaa cattttaaaga cttagttcttggaaaatagctaaata 7080
attatttggt ttgtacagaa taaatatttgcattttaaa tgctttaagaa aagacatccc 7140
atatgtttt gaagtccctcc atattttggaa ataagccaaa tagaaaattt ttattatatt 7200
agcattaaatg ttcaatgctg aaaaaa

<210> 11
 <211> 1691
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7482931CB1

<400> 11
 cgggtcccgcc tgccgagggt taaccccgcc cggtcccggt cctgagctgg accagagccc 60
 tcctccagaa acccctgcgt ccgcccacggc ccaggttaaa tggaaaccac ccttggaaac 120
 tggatgcctg ttagtgcgtt ctaccatatac agtgtattgc aatgagtggg ggaggagagc 180
 agctggatat cctgagtgtt ggaatcttag tgaaagaaaag atggaaagtg ttgagaaaga 240
 ttgggggtgg gggcttggta gaaatttacg atgccttggta catgctcacc agggaaaatg 300
 ttgcactgaa ggtgaatca gctcaacaac caaaacaagt tctgaaaatg gaagttgctg 360
 ttttggaaaaa gctgcaaggg aaagaccatg ttttagatt tattggctgt gggaggaatg 420
 atcgattcaa ctatgtggtc atgcgttgc agggctgaaa tctggcagat ctgcggcgta 480
 ggcggcccg aggacatcc accattaga ccactctccg gctgggtaga cagattttgg 540
 agtctattga aaggattcat tctgtggat tcttgcattcg agacatcaaa ccgtcgaact 600
 tcgtatggg tcgctttctt agtacatgtt ggaatgtta catgcttgc tttggcttgg 660
 ctcgacaatt taccattcc tgggtgacg tcagaccacc tcgagctgtg gcagggtttc 720
 gagggacagt tcgttatgca tcaatcaacg cacatcgaa cagggaaatg ggaagacatg 780
 atgacctttg gtccttattc tacatgttgg tggagttgtt ggttggtcag ctgccttgg 840
 aaaaaataaa ggacaaggag caagtaggct ctattaagga gagatatgac cacaggctca 900
 tggtaaaaca tctccctcca gaattcagca tcttctaga ccatactct tctttggatt 960
 attttacaaa accagactac cagcttctt catccgtgtt tgacaatagc atcaagactt 1020
 ttggagtaat tgagagtgc ccttttgcact gggagaagac tggaaatgtat ggctccctaa 1080
 caaccaccac tacttctacc acccctcaatg tgcacactcg cttgaccctt gctgcaattt 1140
 gaattgccaa tgctactccc atccctggag acttgcttcg agaaaataca gatgaggtat 1200
 ttccagatga acagcttagc gatggagaaa atggcatccc tgggtgtg tcaccagata 1260
 aattgcctgg atctctggc cacccttcgccc cccaggagaa ggatgtttgg gaagagatgg 1320
 atgccaacaa aaacaagatg aagcttggaa tttgttgc tgctactgaa gaggagaaca 1380
 gccatggcca ggcaaatggt cttctcaatg ctccaaggct tgggtccacca attcgtgtcc 1440
 gctcagatg tactcagcca gacagagata ttccactgtt gcgaaatgtt cgttccattt 1500
 acagcttgg gctggaaaaa cgtctgaccc tggagccaaa gccagacact gacaagttcc 1560
 ttgagacctg gtataaaaata gtgttatttt ctttttaaag cttcttaaggt accattatta 1620
 ttgttgcatt tggttattt attattgtat atttctgtt cataaagtct ttcaaataaag 1680
 caaaaaaaaa a 1691

<210> 12
 <211> 6146
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2080788CB1

<400> 12
 cgggctggct gggaaagatgg cggcgggagc ctggggccgccc gcccggccggc cccggccggc 60
 ggagcgaaacc aggggtgtcc ggggtgcgcgt gtccaggggcc gggggccgggc catgagcgcg 120
 ccgtcctcga gtccccggagc cgccggagccc gcccggccccc ctggggccgc cccggcgcccc 180
 tcgccccggc gccggctcgcg gactactcg tgggtgttgc gttcgccgc cccggccgcgc 240
 ggagtggggaa aggccaggcc cagatttgc agcgcttccc agagaaggac tgggaggaca 300
 acccattccc ccaggggcatc gagctgtttt gccagcccgag cgggtggcag ctgtgtcccc 360
 agaggaatcc accgaccttcc tttgttgcgt tcctcaccga catcaactcc gagcgccact 420
 actgcgcctg cttgaccccttcc tgggagccag cggaggcccttcc acaggaaacg acgcgcgtgg 480
 agggatgcaccc agagaggagg gaagaggggg atgagggagg ccagaccac ctgtctcccc 540
 cagcacctgc cccatctgccc cagctgttttgc caccgaagac gctggtaactg gtgtcgccgac 600
 tcgaccacac ggagggtgttcc aggaacagcc ttggcctcat ctatgcacatc cacgtggagg 660
 gcctgaatgt gtgcctggag aacgtgattt ggaacactgtt gacgtgactt gtgcccctgg 720
 ctgggggctc gcagaggacg atctctttgg gggctggta cggcggccatc 780

cactggccga ctcgctgccc gtcagccgct gcagcgtggc cctgcttcc cgccagctag 840
 gcatcaccaa cgtgctgtct ttgttctgtg ccgcctcac ggagcacaag gtctcttcc 900
 tgtcccggag ctaccagcgg ctgcggatg cctgttagggg cctcctggca ctgctttc 960
 ctctcagata cagcttcacc tatgtgcca tcctggccgc tcagctgtg gaggtctca 1020
 gcacaccac gcccctcattt attgggttca acggccctt ccaggcagag acccaggagc 1080
 tgctcgatgt gattttgtct gatctggatg gaggacggt caccatcct gagtgtgtgc 1140
 acattccacc cttgccagag ccactgcaga gtcagacgca cagtgtctg agcatggtc 1200
 tggaccggaa gctggagtt gctgaccccg cttccctcc gcccacgaca tcacacttcc 1260
 ccctgaagat gcagggacaag gagctgcgcg cggcttccct gcccgttcc gctcagctgc 1320
 tgcagggcta tcgctggtgc ctgcacgtcg tgccatcca cccggagcct gtcatccgct 1380
 tccataaggc agcccttcctg ggccagcgtg ggctggtaga ggacgatttc ctgatgaagg 1440
 tgctggaggg catggcctt gctggctt tgtcagagcg tggggtccca taccgcccata 1500
 cggacctgtt cgatgagctg gtggcccacg aggtggcaag gatgcggcg gatgagaacc 1560
 acccccagcg tgtcctgcgt cacgtccagg aactggcaga gcagctctac aagaacgaga 1620
 acccgtaacc agccgtggcg atgcacaagg tacagaggcc cggtgagagc agccacctgc 1680
 gacgggtgcc ccgacccttc ccccggtctgg atgagggcac cgtgcagtgg atcgtggacc 1740
 aggctgcagc caagatgcag ggtgcaccc cagctgtgaa ggccgagagg aggaccaccc 1800
 tgccctcagg gccccccatg actgcccatac tggagcgggtg cagtggctg catgtcaaca 1860
 gcgcccccggc gctggaggtt gtgcgaact gcacccctca cgtttaggg gggaaaatgc 1920
 ttgaggccaa gaagctgtc ccagccgtgt tgagggccct gaaggggcga gctgcccgc 1980
 gctgcctcgc ccaggagctg cacctgcata tgccagcagaa ccgtgcggc ctggaccacc 2040
 agcagtttga ctttgtcgta ctgtatgtgactgctgcctt gcaggactgc acttctctgg 2100
 acgagcatgg cattgcggcg gctctgtcgc ctctggtcac agccttctgc cgaagactga 2160
 gccccggggt gacgcagttt gcatacagct gtgtgcagga gcacgtgtg tgagcacgc 2220
 cacagttctg ggaggccatg ttctatgggg atgtgcagac tcacatccgg gccccttacc 2280
 tggagcccac ggaggacctg gccccccccc aggaggttgg ggaggcacct tcccaggagg 2340
 acgagcgttc tgccctagac gtggcttctg agcagcggcg cttgtgcca actctgagtc 2400
 gtgagaagca gcaggagctg gtgcagaagg aggagagcac ggtgttcagc caggccatcc 2460
 actatgccaa ccgcattgcac tacccctcc tgccccttggc cagcagaag gcccgcctac 2520
 ttcggggagcg tgccgggctg ggcgaccccttggc agagcggccag caacagctg gtcaccaaca 2580
 gcatggctgg cagttgtggcc gagagctatg acacggagag cggcttcgag gatgcagaga 2640
 cctgcgacgt agctggggct gtgttccgc tcatcaaccgc ctttggac aaggtctgca 2700
 cggagagttt ggtcaccagc gaccacccca agggtctgca tgcattgtt ccagacattt 2760
 tccagatgca catcgagacc ctggaggccg tgccagggca gagccggagg ctggccggcc 2820
 tccagaagcc caagctgtg cggccgcgc tgctggccgg tgaggagtgt gtgtggacg 2880
 gcctgcgcgt ctactgtgtc cggatggc gtggaggagg cggggggc agtgcgtggg 2940
 gaccagcatt gctcccagctt gggggcccg tcttcctcac cacgtaccgg gtcatttc 3000
 cggggatgcc cacggaccccc ctggttggg agcaggttgg ggtccgtcc tttccgggtgg 3060
 ctgcgctgac caaggagaag cgcatcagcg tccagacccca tgcattgtt ccctgtcagg 3120
 acgggctcca gtcgcgttcc tgcacatttca agtgcgtgaa aatggcttt gacgaggagg 3180
 tggggtctga cagcggcgag ctttccgttca agcagctgca taagctgcgg taccggccgg 3240
 acatcagggc cacccttgcg ttcaccttgg gctctgccc cacacctggc cggccaccgc 3300
 gagtcaccaa ggacaagggtt cttccctca gaaccctgtc ccggaaactg gtcaagaacg 3360
 ccaagaagac catcgccggc cagcatgtca ctgcgaagaa gtacaacccc cccagctggg 3420
 agcaccgggg ccagccgccc cctgaggacc aggaggacga gatctcagtgc tcggaggagc 3480
 tggagcccaag cacgtgacc ccgtccctcg ccctgaagacc ctccgaccgc atgaccatga 3540
 gcagcctgtt gggaaagggtt tgctgtcgcc actaccagcg cctcggctcg ggacccctga 3600
 gcagcagctt gaggccggcc aagtctgacg cctccgcattt ctctccggc aaccgcattt 3660
 atgcacatcg ccgcacgtac ccagggttgc tgatctgtcc ccagagtgtc caggacaacg 3720
 ccctgcagcg cgtgtcccgcc tgcttccgcg agaaccgtt cccctggc tgcggccga 3780
 gccccgggtt caaggcggtt ctgtgcgtt ctggaggccct gcatggaaa ggtgtcgctg 3840
 gcctcttcaa ggccagaac gcaccccttc cagggccagtc ccaggccggac tcgagtagcc 3900
 tggagcagga gaagtacctg caggctgtgg tcaacttccat gccccgttcc gccgacgcgt 3960
 cgggacgcaa cacgtttagc ggcttcttcc cagccccatcat gggcagtcac ggttaagtggg 4020
 gcagtgctcg gaccagtggc cgcagcgtg gccttggcac cgatgtggc tcccggttag 4080
 ctggcagaga cgcgcgtggcc ccaccccaagg ccaacgggggg ccctccgcac cccggcttcc 4140
 tgcgtccgca gcgagcagcc ctctatatacc ttggggacaa agcccagctc aagggtgtgc 4200
 ggtcagaccc cctgcagcg tgggagctgg tgccatttga ggtattcgag gcacggcagg 4260
 tgaaggctag cttcaagaag ctgtgttttca catgtgtccc aggctggccc gtcgtcgac 4320
 ccagcccaagc ctcccttcctg cgctcactgg aggactcaga gtggctgatc cagatccaca 4380
 agctgctgca ggtgtctgtt ctgggttgg agtctcttgg ttcaggctcc tccgtgttcc 4440
 tgggcctgaa ggatggctgg gacatcacca cccaggttggt atccctggg tcaactgtct 4500
 cagacccctt ctaccgcacg ctggagggtt ttcgcctgtt ggtggagaag gagtggctgt 4560

ccttcggcca	tcgcgttcagc	caccgtggag	ctcacaccct	ggccgggcag	agcagcggct	4620
tcacacccgt	cttcctgcag	ttcctggact	gcgtacacca	ggtccacctg	cagtccccca	4680
tggagtttga	gttcagccag	ttctcacctca	agttcctcgg	ctaccacca	gtgtccccgc	4740
gtttccggac	cttcctgctc	gactctgact	atgagcgcac	tgagctgggg	ctgctgtatg	4800
aggagaaggg	ggaacgcagg	ggccagggtc	cgtcagggtc	tgtgtggag	tatgtggacc	4860
ggctgagcaa	gaggacgcct	gtgttccaca	attacatgt	tgcggccgag	gacgcagagg	4920
tcctcgccgc	ctacagcaac	gtgtccaacc	tgaagggtgt	ggacttctac	actgaggaga	4980
cgctggccga	gggcctccc	tatgactggg	aactggccca	ggggccccc	gaacccccag	5040
aggaagaacg	gtctgtatgga	ggcgctcccc	agagcaggcg	ccgcgtgtg	tggccctgtt	5100
acgacagctg	cccgccggcc	cagcctgacg	ccatctcacg	cctgctggag	gagctgcaga	5160
ggctggagac	agagttgggc	caaccgcgt	agcgctggaa	ggacacctgg	gaccgggtga	5220
aggctgcaca	gcmcctcgag	ggccggccag	acggccgtgg	cacccttagc	tccctccttg	5280
tgtccaccgc	acccaccac	cgtcgctcgc	tgggtgtgt	cctgcaggag	ggggccgtgg	5340
gctccaccct	gagcctcagc	ctggacagcg	accagagtag	tggctcaacc	acatccggct	5400
cccgctcaggc	tgcccgccgc	agcaccagca	ccctgtacag	ccagttccag	acagcagaga	5460
gtgagaacag	gtcctacgag	ggcaactctgt	acaagaaggg	ggccttcatg	aagccttgg	5520
aggcccgctg	gttcgtgtcg	gacaagacca	agcaccagct	gcgcctactac	gaccacccgt	5580
tggacacaga	gtgaagggt	gtcatcgact	tggcgaggt	ggaggctgt	gcacctggca	5640
cggccactct	gggtggccct	aagaccgtgg	acgagaaggc	cttcttgac	gtgaagacaa	5700
cgcgtcgct	ttacaacttc	tgtgcccagg	acgtgcctc	ggcccaagcag	tgggtggacc	5760
gatccagag	ctgcctgtcg	gacgcctgag	cctcccagcc	ctgccccggct	gctctgttc	5820
cggtcgttac	cgaccactag	gggtggccag	ggccgccccg	gccatgtta	cagccccggc	5880
cctcgacagt	attgaggccc	cgagccccc	gcacttgtgt	gtacagcccc	cgtccccggc	5940
ccgccccggc	cgccggggcc	taacttattt	tggcgtcaca	gctgagcacc	gtgcccgggag	6000
gtggccaagg	tacagccccg	aatgggcctg	taaatagtcc	ggcccccgtca	gcgtgtgt	6060
gtccagccag	cggctgcagg	cgagtttcta	gaaccagagt	ctatataaag	agagaactaa	6120
cgccaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	6146

<210> 13
<211> 2362
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71918969CB1

<400> 13

gctctcagcgt	gtgaatatgt	gtcttagtgca	tccttaacct	gaggacttca	ccagttcgaa	60
attacagttt	tcaccatcaa	ctacattatc	ctttttggcc	tggtttctt	cctcaaacag	120
tggaaacatt	tttaaagggt	cttttgtgc	agagttaaac	aaatggctga	tagtggctt	180
gataaaaaat	ccacaaaatg	ccccgactgt	tcatctgctt	ctcagaaaaga	tgtactttgt	240
gtatgttcca	gcaaaaacaag	ggttcctcca	gttttgggt	tggaaatgtc	acagacatca	300
agcattggta	gtgcagaatc	ttaatttca	ctggagagaa	aaaaagaaaa	aaatatcaac	360
agagatataa	cctccaggaa	agatttgc	tcaagaacct	caaagttaga	gagaaaagca	420
tctcagcaac	aatggggctg	ggcaactt	acagaaggaa	aagttcctca	cataaggatt	480
gagaatggag	ctgcattatg	ggaaatctat	acctttggaa	aatatttgg	aaaaggggagc	540
tttggaaatag	tcatttgaagc	gacagacaaag	gaaacagaaa	cgaagtgggc	aattaaaaaa	600
gtgaacaaaag	aaaaggctgg	aagctctgt	gtgaagttac	ttgaacgaga	gttgaacattt	660
ctgaaaatgt	taaaacatg	acacatcata	catttggaa	aagtatttga	aacgcacaaag	720
aaaatgtacc	ttgttgcgtt	gttttgcgtt	gatggagaac	tcaaaggaaat	tctggatagg	780
aaaggccatt	tctcagagaa	tgagacaagg	tggatcattc	aaagtctcgc	atcagctata	840
gcataatctc	acaataatg	tattgtacat	agagatctg	aactggaaaa	tataatgggt	900
aaaagcagtc	ttatttgcgt	taacaatgaa	ataaactttaa	acataaagg	gactgatttt	960
ggcttagccg	tgaagaagca	aagttaggat	gaagccatgc	tgcagggccac	atgtgggact	1020
cctatctata	tggcccccgt	agttatcgt	gcccacgact	atagccagca	gtgtgacatt	1080
tggagcatac	cgctcgtaat	gtacatgtt	ttacgtggag	aaccaccc	tttggcaagg	1140
tcagaagaga	agcttttgc	gttaataaga	aaaggagaac	tacatttga	aatgcagtc	1200
tggaaattcca	taagtgcactg	tgctaaaatg	gttttggaaac	aacttatgaa	agtagatcct	1260
gctcacagaa	tcacagctaa	ggaactacta	gataaccagt	ggttaacagg	caataaaactt	1320
tcttcgggt	gaccaaccaa	tgttattagag	atgatgaaagg	aatggaaaaa	taacccagaa	1380
agtgttggagg	aaaacacaac	agaagagaag	aataagccgt	ccactgaaga	aaagttgaaa	1440
agttaccaac	cctggggaaa	tgtccctgtat	gccattaca	cttcagatga	agaggaggaa	1500

aaacagtcta ctgcttatga aaagcaattt cctgcaacca gtaaggacaa ctttgatatg 1560
tgcagttcaa gtttcacatc tagcaaactc cttccagctg aaatcaaggg agaaatggag 1620
aaaacccctg tgactccaag ccaaggaaca gcaaccaagt accctgctaa atccggcgcc 1680
ctgtccagaa ccaaaaagaa actctaaggt tccctccagt gttggacagt acaaaaacaa 1740
agctgctctt gtttagcatt ttagtgggg gtaggagggg aagaagacag ccctatgctg 1800
agctttagtc cacagagccc cgccatgtgt ttgcaccagc ttaaaaattga 1860
agctgcttat ctccaaagca gcataagctg cacgtggcat taaaggacag ccaccagtag 1920
gcttggcagt gggctgcagt ggaaatcaac tcaagatgtc cacgaaggtt ttttaggggg 1980
gcagataacct tcaatttaag gctgtggca cacttgctc ttttacttc aaattcttat 2040
gtttaggcac agctatttat aggggaaaac aagaggccaa atatagtaat ggaggtgcca 2100
aataattatg tgcactttgc actagaagac ttttttagaa aattactaat aaacttgcca 2160
tacgtattac agcagaagtg cttcagtcat tcacatgtgt tcgttagatt ttaggttgct 2220
atagattgtt taagacagct tattttaaat gtagaaaaat aggagattt gtaactgctt 2280
gccattaact tgctgctaaa ttcccaatgt attgattaaa tcaatcaaaa aacagatgtt 2340
cctcagcaca aacacacacaa aa 2362

<210> 14
<211> 1535
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8187571CB1

<400> 14
ccaccaccc ccttagcctc ctgtgtggga ggagttatg ggtatgtggc tcctgcccag 60
tccaggtgg ctttcacttc tactctattt cagttccctt ttcccgtatct gggctggaga 120
gcttcctat tgtaaggca gcagaaacct tcgcgtggatg gtttttaggat aaggggtcat 180
caatgctggc aagagtcggc acaatgagga ccaggcttgc tgtgaagtgg tttatgtgg 240
aggtcggagg agtgcgtacag ggttacccatgg gggcgttgc cgaggccagg gactctgtt 300
ctactactgg ggccttattt atggggcatgc agggggcgga gctgctgaaa tgccctcacc 360
gctcctgcat cgccatatacc gagagcagct aaggacatgc gtagagatac ttccaggacc 420
ttccgcacca cccctctgcc tcccaaccac tccggggacc ccagatccctt ccgcattcc 480
tcacttgctt ggcctcagt cctgctgtc ttccacagaag gaagttagcc acgagagcct 540
gttagtgggg gccattgaga atgccttcca gctcatggat gagcagatgg cccgggagcg 600
gcgtggccac caagtggagg ggggctgtc tgcaactgggtt gtatctacc tgcttaggca 660
gtgtacgtg gccaatgcag gcgatagcag ggcatttgcgtt gtcggaaatg tgaaatcat 720
tccaatgtcc cgggagttt ccccgagac tgagcgtccag cgttccatc tgctggctt 780
cctgaaacca gagctgctag gcagtgaatt caccacattt gagttcccc gcagagttct 840
gcccaaggag ctggggcaga ggttgcgtt ccgggaccag aacatgaccg gctgggccta 900
caaaaaagatc gagctggagg atctcaggtt tcctctggc tttggggagg gcaaaaaaggc 960
tcgggtgtatg gccaccattt gggtgaccgg aggttgggg gaccacagcc ttaagggtctg 1020
cagttccacc ctgcccattca agcccttctt ctctgttcc cctgaggtac gagtgtatga 1080
cctgacacaa tatgagact gcccagatga tttgtgttgc ctgggaacag atggcctgtg 1140
ggatgtact actgactgtg aggttagctgc cactgtggac aggggtgtgt cgccctatga 1200
gcctaatgac cacagcaggat atacagctt ggccttgcgtt ctgggtccgg gggcccccggg 1260
taccccccga gaccgtggct ggcgttccca caacaacaag ctgggttccg gggatgacat 1320
ctctgtctt gtcattttttt tggggggcc aggcgttac tcctgagggg ctgaacaccca 1380
tccctccac tagctctcc atacttactc ctctcacagc ccaaattctg aagttgtctc 1440
cctgaccctt cttagtggc aacttaactg aagaaggat gtccgctata tccaaaattta 1500
cagctattgg caaataaaacg agatggataa aaaaa 1535

<210> 15
<211> 1376
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7494145CB1

<400> 15

<210> 16
<211> 1482
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 5807954CB1